AN INVESTIGATION INTO THE EXIT LEVEL ASSESSMENT OF THE CLINICAL COMPETENCE OF MEDICAL GRADUATES FOR THE SUB-SAHARAN AFRICAN CONTEXT

by

Christina Phoay Lay Tan

Dissertation presented for the degree of
Doctor in Philosophy
in the
Faculty of Medicine and Health Sciences
at
Stellenbosch University

Supervisor: Prof. Susan C van Schalkwyk

Co-supervisor: Prof. François J Cilliers

December 2017

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2017

Christina Phoay Lay Tan

SUMMARY

Assessment plays a major role in the process of medical education and is an essential component of any medical curriculum. In order to allow judgements to be made on the competence of students, the quality of assessment methods needs to be assured. As exit-level assessments are used for licensing and certification purposes, they are regarded as "high-stakes" in nature, with significant implications for the student, the curriculum, the institution and the public.

This study investigated how valid exit-level assessment practices are in determining the clinical competence of medical graduates for selected health-related issues, with the focus on selected medical schools in Sub-Saharan Africa where there are resource constraints and a high burden of disease. It was warranted in light of the paucity of published information about assessment practices and factors that could influence the decisions that are made in exit level assessment in Sub-Saharan Africa

The study was conducted within a qualitative interpretive paradigm. A range of data-gathering methods were utilised. The research design selected was multiple case studies, with the unit of analysis the exit-level assessment practices of undergraduate medical curricula. Three medical schools in Sub-Saharan Africa were purposively selected, and as a proxy for the entire curriculum, the clinical disciplines of Obstetrics & Gynaecology and Paediatrics were specifically chosen. Data sources included document analysis, the observation of exit level assessments, and individual and group interviews with faculty involved in exit level assessments. The audio-recorded interviews were transcribed and inductive thematic analysis was carried out. Data from documents were mapped on Excel spreadsheets and an observation schedule was used to collect data during observation of actual clinical assessments. Two analytical tools, Miller's 'pyramid' (1990) and Disability Adjusted Life Years (DALYs), were then used for deductive analysis of data obtained from documents and observations.

There were a number of findings of note. A selection of assessment methods were used at the three study sites. Although assessment question content was appropriate and focussed on relevant contextual health-related issues, the evidence gathered through current assessment practices is insufficient to allow accurate judgements to be made about the clinical competence of the students. DALYs as an analytical tool has the potential to contribute to a process of blueprinting, but its limitations in its use with the non-disease based curriculum have been highlighted. Examiners were key players involved in exit-level assessment and have considerable influence on assessment practices. However, their lack of understanding of the basic concept of assessment suggests a need for faculty development to raise the level of assessment literacy. The context within which assessment occurs and institutional cultures also influences assessment practices.

This study has provided insights into exit level assessment practices taking place in a resource-constrained context such as Sub-Saharan Africa. It is envisaged that this research will add to the body of knowledge and that it will inform assessment practices going forward, as well as potentially informing curriculum design.

OPSOMMING

Assessering speel ’n uiters belangrike rol gedurende mediese onderrig en is ’n noodsaaklike komponent van enige mediese kurrikulum. Ten einde uitspraak te kan lewer oor die vaardigheid van studente, is dit noodsaaklik dat die kwaliteit van die assesseringsmetodes verseker word. Aangesien uittreevlak-assesserings gebruik word vir lisensiërings- en sertifiseringsdoeleindes, word hulle as inherent uiters belangrik beskou, met belangrike gevolge vir die student, die kurrikulum, die inrigting en die gemeenskap.

Hierdie studie het ondersoek ingestel na die geldigheid van uittreevlak-assesseringspraktyke in die vasstelling van die kliniese vaardigheid van mediese graduandi vir spesifieke gesondheidsverwante probleme. Die fokus is op geselekteerde mediese skole in sub-Sahara Afrika waar daar beperkte hulpbronne en ’n hoë siektelas is. Die studie is geregverdig in die lig van die gebrek aan gepubliseerde inligting oor assesseringspraktyke en faktore wat die besluite kan beïnvloed wat geneem word in uittreevlak-assesserings in sub-Sahara Afrika.

Die studie is uitgevoer binne 'n kwalitatiewe interpretatiewe paradigma. ’n Wye verskeidenheid data-insamelingsmetodes is gebruik. Die navorsingsontwerp wat gekies is, was veelvuldige gevalle-studies. Die analise-eenheid was die uittreevlak-assesseringspraktyke van voorgraadse mediese kurrikula. Drie mediese skole in sub-Sahara Afrika is doelbewus gekies, terwyl die kliniese dissiplines van Verloskunde & Ginekologie, en Pediatrie gekies is as verteenwoordigend van die hele kurrikulum. Databronne sluit in: dokument-analise; waarneming van uittreevlak-assesserings, asook individuele en fokusgroep-onderhoude met die betrokke fakulteite in die uittreevlak-assesserings. Oudio-opnames van onderhoude is getranskribeer, waarna induktiewe tematiese analise gevolg het. Data uit dokumente is op Excel-sigblaaie aangeteken en ’n waarnemingskedule is gebruik om data te versamel tydens die waarneming van werklike kliniese assesserings. Twee analitiese instrumente, naamlik Miller se “piramide” (1990) en *Disability Adjusted Life Years* (DALYs)/(Ongeskiktheids-

aanpassende lewensjare), is gebruik vir die deduktiewe analise van die data wat verkry is uit dokumente en waarnemings.

Daar was ’n aantal belangwekkende bevindinge. ’n Seleksie van assesseringsmetodes is gebruik by die drie mediese skole waar die studie plaasgevind het. Alhoewel die inhoud van die assesseringsvrae gepas was en gefokus het op relevante, kontekstuele gesondheidsverwante sake, is die inligting wat deur die huidige assesseringspraktyke verkry is, onvoldoende om ’n akkurate assessering van die kliniese vaardighede van die studente te maak. DALYs het die potensiaal as ’n analitiese instrument om by te dra tot die proses om ’n bloudruk te verkry, maar die tekortkoming in die gebruik daarvan in die nie-siekte verwante komponente van die kurrikulum is beklemtoon. Eksaminatore is sleutel-rolspelers wat betrokke is by uittreevlak-assesserings en het ’n beduidende invloed op assesseringspraktyke. Die gebrek aan begrip van die basiese konsep van assessering dui egter op die behoefte aan fakulteitsontwikkeling om die vlak van assesserings-geletterdheid te verhoog. Die konteks waarbinne assessering plaasvind en institusionele kulture beïnvloed ook assesseringspraktyke.

Hierdie studie het insig gegee in die uittreevlak-assesseringspraktyke wat in sub-Sahara Afrika, met sy beperkte hulpbronne, plaasvind. Daar word in die vooruitsig gestel dat hierdie navorsing sal bydra tot die kennisbasis wat assesseringspraktyke, en moontlik ook kurrikulum-ontwerp, in die toekoms sal toelig.

ACKNOWLEDGEMENTS

I am truly indebted to a great number of people who I have had the privilege of engaging with in different corners of the world at various times in my life. Some of you planted the seeds of interest in medical education a long time ago that have germinated and taken firm root. Some of you have helped me in one way or another on this doctoral journey of discovery, as mentors, role models, friends and supporters, colleagues, and family. Without your guidance, support and understanding, this would not have been possible at all. To all of you, thank you.

To my parents, Peter and Eleanor, thank you for being my biggest inspiration in life and for your endless love. I hope I have fulfilled the aspirations and dreams that you yourselves may have had but did not have the opportunity to pursue. And to Francis, Irene, Patrick, Lucy, and all the family: thank you for all your strong steady support and patience.

To my supervisor, Professor Susan van Schalkwyk, and my co-supervisor, Professor François Cilliers, it has been a real privilege to have you both as my *sifu*. You had more confidence in my abilities than I did, and through tough love you have taught me how to stand tall with hands on hips. This journey really would not have been achievable without your insights, and help and guidance all the way, especially when times were challenging. To both of you, my most heartfelt thanks.

This PhD would not have been possible without financial assistance from the Stellenbosch University Rural Medical Education Partnership Initiative (SURMEPI) and from the Foundation for the Advancement of International Medical Education and Research (FAIMER). I am also very grateful to the faculty, students and interns who participated in the study. Thank you all for your time and help.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Overview of the Study

1.1 Introduction and Background

Assessment plays a major role in the process of medical education (Shumway and Harden 2003) and is an essential component of any medical curriculum. It has been described as "any purported or formal action to obtain information about the competence and performance of a candidate" (Schuwirth and van der Vleuten, 2010). The assessment of clinical competence is one of the most important tasks facing medical teachers and usually occurs periodically during, and at the exit point from, a degree to certify a level of achievement (Newble, 1992; Wass et al., 2001c; Epstein, 2007). A range of methods is typically used to assess clinical competence (Wass et al., 2001b; Barzansky and Etzel, 2003; Howe et al., 2004; Wilkinson and Frampton, 2004; Roberts et al., 2006; van der Vleuten et al., 2010).

In order to allow judgements to be made on the competence of students, the quality of the assessment methods needs to be assured. This includes designing assessments to provide reliable and valid measurements of student performance (Wass et al., 2001c; Epstein, 2007; Wilkinson, 2007). Validity in the context of assessment in medical education refers to the evidence presented to support or refute the meaning or interpretation assigned to assessment results (Downing, 2003). Reliability denotes reproducibility of assessment data or scores, over time or occasions (Downing, 2004:1006).

Understanding what the term "competence" means in this context is key to ensuring that assessment decisions are valid. "Competence" broadly speaking is the "ability to do something well" or could equally be a "skill that you need in a particular job or for a particular task" (Oxford University Press, 2015). In the literature, the definition varies depending on the context, and is used in fields as diverse as education, linguistics, biology, psychology, management, human resources, and politics (Hager and Gonczi,

1996; Hoffmann, 1999; Taylor, 1988; Le Deist and Winterton, 2005; Jeris et al., 2005). Competence in medicine has been defined as "the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served" (Epstein and Hundert, 2002:226). This poses challenges for assessment. Individual skills or attributes can be measured in isolation, and perhaps more easily as discrete tasks. However, the intricacies of assessing competence are a complex, multidimensional undertaking (Miller, 1990; Carraccio et al., 2002; Frank et al., 2010a). Appropriate assessment to measure competence in medical graduates, therefore, remains a challenge.

Medical schools have a responsibility to ensure and demonstrate that a certain level of competence has been achieved in their graduating doctors (Shumway and Harden, 2003), as this is central to public accountability. They are also increasingly being challenged to provide evidence that the assessments used are discriminating between the sufficiently and "minimally competent" (Boursicot and Roberts, 2006) at all levels of medical training and education (General Medical Council, 2009; General Medical Council, 2011; Liaison Committee on Medical Education, 2013). As exit-level assessments are used for licensing and certification purposes (Epstein, 2007), they are regarded as "high-stakes" in nature, with significant implications for the student, the curriculum, the institution and the public (Epstein and Hundert, 2002; Roberts et al., 2006). Although the existing literature in support of issues on assessment and competence is extensive, questions arise regarding the extent to which they can be generalised and applicable across different contexts, whether local factors influence assessment or determine what competence is relevant to assess.

1.2 Context of the study

This study investigated how valid exit-level assessment practices are in determining the clinical competence of medical graduates for selected health-related issues, with the focus on selected medical schools in Sub-Saharan Africa where there are resource constraints and a high burden of disease. Historically, medical training programmes or courses in this setting were established in accordance with the traditions of colonial occupiers predominantly in the 19th and first half of the 20th centuries. Western curricula were imported, along with assessment methods, for use in African medical

schools in order to achieve comparable standards in training to those found in the West (Gukas, 2007). It could be argued that while the current training and assessment of students at Sub-Saharan African medical schools may provide appropriately competent graduates by such criteria, there appears to be no supporting evidence at present as to whether this assessment is equally valid for the resource-constrained settings that they will be practising in. Collecting such evidence in support of decisions about competence in a Sub-Saharan African context poses two challenges. The first is what graduates should be certified as being competent for. The second, and more important challenge, is collecting sufficient and appropriate evidence – in a context where resources are scarce – to support those decisions about competence that were made at the time of student assessment.

In resource constrained settings, these constraints include serious financial limitations and lack of manpower (Walubo et al., 2003; Mullan et al., 2011; Greysen et al., 2011; Chen et al., 2012; Organisation for Economic Co-operation and Development, 2013). Such constraints affect not only the setting of clinical service delivery where health-care professionals have to contend with dealing with the high burden of disease but also the medical education setting where medical schools rely on health-care professionals from the clinical service context to teach. Exit-level assessment takes place at the interface of clinical service and medical education. In such a context, there are challenges to aligning assessment practice with theoretical ideals (see, for example, Norcini et al. 2011). Allocating adequate resources to collecting assessment-related evidence – and therefore allowing valid assessment-related decisions – also poses a challenge.

The 2006 WHO World Health Report, which focused on health workforce issues, recorded that Africa had just 1% of the total global funding for health, and only 2.3 health-care workers per 1000 population (compared with 18.9 in Europe) dealing with 24% of the global disease burden (World Health Organisation, 2006). The lack of doctors coupled with high clinical workloads and working in academic institutions where teaching does not have the same status or priority as research (Frenk et al., 2010) could impact significantly on the teaching that takes place, what is assessed and how the assessment is conducted.

Regarding what graduates in Sub-Saharan Africa should be certified as competent for, Frenk et al. (2010) have made an impassioned plea for changes in the education of

health professionals in order to provide graduates who are better equipped to deal with the challenges of health-care in the 21st century. The mismatch of competencies to patient and population requirements (Frenk et al., 2010) needs to be addressed, a sentiment echoed by Miller et al. (2011) when proposing the alignment of health sciences education programmes with the health-care needs of a particular country. In aligning the education of health professionals with the local health-care needs of the community, it would also follow logically that assessment needs to be aligned with the curriculum.

Defining the health-care needs of a region like Sub-Saharan Africa can be complex. One example of how health-care needs and priorities have been delineated is through the Millennium Development Goals. The United Nations Millennium Declaration (United Nations, 2000), signed in September 2000 by all 189 UN member states, committed world leaders to combat poverty, hunger, disease, illiteracy, environmental degradation, and discrimination against women. Eight Millennium Development Goals (MDGs) were derived from this Declaration (World Health Organisation, 2012). Of these, three are of direct relevance to defining appropriate competence for doctors in Sub-Saharan Africa, namely:

MDG 4: To reduce child mortality

MDG 5: To improve maternal health

MDG 6: To combat HIV/AIDS, malaria and other diseases.

Another example of defining health-care needs in Sub-Saharan Africa is through the Global Burden of Disease study, initiated in 1992, which sought to quantify the global burden of premature death, disease, and injury with measures that could be used for cost-effectiveness analysis (Murray and Lopez, 1997). One such measure used is the disability-adjusted life years (DALYs), which allows for the comparison of the burden of premature mortality with non-fatal health outcomes, such as disability (World Health Organisation, 2015). However the health-care needs are defined, health professionals, including doctors, would need to have appropriate training and achieve a level of competence in order to be able to meet these health related issues.

Given the context as outlined previously, important questions arise: (1) What exit-level assessments take place in Sub-Saharan African medical schools, where there are

significant resource constraints, and (2) Are these exit-level assessments valid measures of the clinical competence required of Sub-Saharan African medical graduates as defined by contextual health-related issues?

Currently, there are only a small number of studies in the literature regarding clinical assessment of medical students in the Sub-Saharan African context. Burch and Seggie (2008) reported on the use of portfolio assessment using a structured interview in a resource-constrained setting, which has since been used in other institutions (Cameron et al., 2011; van Schalkwyk et al., 2012). The Sub-Saharan African Medical Schools (SAMSS) Study provided baseline data on the medical schools including infrastructure and capacity (Mullan et al., 2011; Chen et al., 2012) as well as a review of existing literature on medical education in Sub-Saharan Africa (Greysen et al., 2011). These reports do not address current assessment practices in medical schools on the continent, nor do they describe what measures are taken in these institutions to ensure that these assessments are appropriate in this context.

Walubo et al. (2003) developed a hypothetical model and proposed guidelines on the selection of sustainable assessment methods in the limited resource settings of most African medical schools. Their guidelines scored and weighted selected assessment methods against performance, cost, suitability and safety factors. Using their formula, they determined that a problem-based oral examination fared better than the Objective Structured Clinical Examination (OSCE) or a patient-based clinical examination, and essays and short answer questions (SAQs) were more appropriate than multiple choice questions (MCQs) as choice of assessment methods in an African medical school setting. Walubo et al.'s model remains largely a hypothetical one as there does not appear to be any published literature regarding its application in any medical school in Sub-Saharan Africa. However, it has at least provided some consideration for assessment methods for resource-constrained settings.

In light of this scarcity of information about clinical assessment, this lead on to the question: What current assessment practices are actually taking place in Sub-Saharan African medical schools? A starting point was to determine what assessment was presently taking place in one Sub-Saharan African medical school.

1.2.1 Preliminary study

A preliminary study was carried out in April and May 2013, to map current assessment practice in the final 18-month phase of the six-year Bachelor of Medicine & Bachelor of Surgery (MB,ChB) programme at Stellenbosch University, based on the way in which assessment methods are described in the relevant official documentation (Tan et al., 2016). Document analysis of the final phase (2012/2013) study guides was undertaken, in conjunction with interviews with clinical-clerkship chairs to verify that the content of the study guides was correct. Analysis of the data obtained was done using an existing framework for the assessment of clinical competence (Miller, 1990), also known as Miller's 'pyramid'.

The findings from this preliminary study called into question the strength of decisions made at exit-level about medical graduates and highlighted potential areas where current practice needed to be investigated in greater depth (Tan et al., 2016). There was lack of uniformity in how assessments were described and also in the weightings (relating to the calculation of the students' final marks). Similarly, the issue of available resources in relation to assessment methods used has surfaced. This study therefore provided the impetus for further investigation into exit-level assessment in a resource-constrained setting.

1.3 Research Aim and Objectives

The research presented in this thesis thus aimed to investigate what current (that is, at the time of this study) exit-level assessment practices in selected Sub-Saharan African medical schools revealed about apparent assumptions regarding the clinical competence required of medical graduates in resource-constrained settings to respond to selected health-related issues. The research question was, "How valid are the exit-level assessment (ELA) practices in selected medical schools in Africa in determining the clinical competence of medical graduates for selected health issues in the Sub-Saharan African context?"

The study had the following objectives:

- To investigate what ELA methods were currently stated as being used in resource-constrained settings

- To investigate what ELA methods were currently actually being used in resource-constrained settings
- To explore the reasons for the choices of these ELA methods, and
- To investigate what clinical competencies were considered appropriate for medical graduates from these medical schools in order to address selected health-related issues.

The analysis of data generated from the study in answer to the above objectives would then contribute to answering the research question as stated previously.

1.4 Overview of Research Methodology

1.4.1 Research Design

This study was conducted within a qualitative interpretive paradigm. Qualitative data-gathering methods were utilised and the research design selected for this study was multiple case studies. The unit of analysis was exit-level assessment practices of undergraduate medical curricula. The choices made and rationale behind these decisions will be elaborated on further in Chapter 4.

1.4.2 Cases and Sampling

The study took place between April 2014 and June 2015 at three medical schools situated in three different Sub-Saharan African countries, namely, (1) the University of Botswana School of Medicine in Botswana, (2) Stellenbosch University Faculty of Medicine and Health Sciences in South Africa, and (3) Makerere University College of Health Sciences in Uganda. These schools were purposively selected because they spanned a wide range in the number of years they had been established. These three institutions were also Medical Education Partnership Initiative (MEPI) grant sites. This US-funded five-year initiative (2010-2015) aimed to support medical education and research in Sub-Saharan African institutions, with a focus on strengthening manpower resources in order to address national health needs (Medical Education Partnership Initiative, 2013). As this research study was in line with one of MEPI's core themes emphasizing regionally relevant research, it appeared reasonable to consider conducting the study in selected MEPI grant sites. A further reason for the choice of these institutions was that one language common to all three settings was English, as at

the time of planning it was envisaged that the fieldwork and data collection at these three sites would predominantly be carried out in this language.

Faculty members, including clinical-clerkship coordinators (or discipline chairs), from the Departments of Paediatrics & Child Health and Obstetrics & Gynaecology in the three medical schools who were involved in exit-level assessment were invited to participate in the study.

Interns from Stellenbosch University and Makerere University were also invited to participate in the study. The first cohort of students from the University of Botswana graduated only in the latter half of 2014, so there were no interns available to participate in this study at the time of application of ethical approval for the study.

1.4.3 Data Collection

Various methods of data collection were utilised in each of the three cases, in relation to the study objectives.

- ***To investigate what stated ELA methods were currently stated as being used*:**
  (a) Document analysis of all documents relating to exit-level assessments was carried out. This analysis provided information on current ELAs.
  (b) Clarificatory interviews with clinical-clerkship coordinators (or discipline chairs) were conducted, solely for the purpose of verifying the content of the study guides and other ELA documentation was correct.

- ***To investigate what actual ELA methods were currently actually being used*:**

  Observation of ELAs at the respective medical schools was also undertaken, where permissible, using an observation schedule.

- ***To explore the reasons for the choices of Exit-Level Assessment methods*:**
  (a) Semi-structured individual interviews (DiCicco-Bloom and Crabtree, 2006) were conducted with clinical-clerkship coordinators (or discipline chairs) from the Departments of Paediatrics & Child Health and Obstetrics & Gynaecology in the respective medical schools involved in exit-level assessment.

(b) Group interviews (DiCicco-Bloom and Crabtree, 2006) were conducted with faculty from the Departments of Paediatrics & Child Health and Obstetrics & Gynaecology in the respective medical schools involved in exit-level assessment.

- ***To investigate what clinical competencies were considered appropriate for medical graduates from these selected medical schools in order to address selected health-related issues:***

  (a) The semi-structured individual interviews with clinical-clerkship coordinators (or discipline chairs) as described above informed this research objective.

  (b) Similarly, the group interviews with faculty involved in exit-level assessment as described above also informed this research objective.

  (c) Interns (who graduated within the last 1-2 years) from two of the three medical schools were invited to participate in group interviews.

- ***To investigate the validity of Exit-Level Assessments in determining the clinical competence of Sub-Saharan African medical school graduates to address selected health-related issues*:**

  A response to this final research objective was formulated from all the data that gathered and subjected to analysis, interpretation and synthesis.

1.4.4 Data Analysis

Data obtained in the form of digital-audio recordings made during the interviews was transcribed with the assistance of a professional transcribing service and transcripts of the interviews were also sent back to the participants for data checking. Inductive thematic analysis (Elo and Kyngäs, 2008; Tavakol and Sandars, 2014b) was used for analysing the data with careful reading and re-reading of the text of transcriptions to help identify and index themes and categories. All data relevant to each category was identified and examined by constant comparison. A computer software programme was used to assist with the qualitative data analysis.

Analysis of the previously-mentioned ELA documents was mapped using an Excel spreadsheet to provide an overall view and to facilitate comparison of the ELAs between the three medical schools.

The issue of whether undergraduate training in Sub-Saharan African medical schools enabled graduates to address contextual health-related issues was mentioned earlier. Millennium Development Goals (MDGs) 4 and 5 related directly to the disciplines of Paediatrics (or Child Health) and Obstetrics and Gynaecology (or Maternal Health) respectively in clinical medicine. As a starting point in this study, it was plausible to focus on MDGs 4 and 5, and therefore, the disciplines of Paediatrics and Obstetrics & Gynaecology, in investigating the validity of the exit-level assessments and the clinical competence of medical graduates. However, in the course of data analysis, it became evident that while the MDGs provided a broader overview to inform health policies, they were at too general a level to inform medical curricula. After further investigation, the decision was taken to use Disability Adjusted Life Years (DALYs) to appraise the alignment of assessment in medical schools with pertinent local health-care issues and societal need.

1.4.5 Ethical Considerations

Institutional ethical approval was sought and obtained from each medical school in turn to conduct this research study. Faculty from the Departments of Paediatrics & Child Health and Obstetrics & Gynaecology as well as interns (recent graduates) of the respective medical schools were invited to participate in the study. Participation was voluntary and written informed consent was obtained from each participant. In all instances of data collection, the principles of ethical practice in conducting research were practised and adhered to (these will all be described in detail in Chapter 4).

1.5 Structure of the thesis

Chapter 1 has provided a brief overview of this study, including the rationale, research aim and objectives and the research methodology. In Chapter 2, theoretical perspectives underpinning the study, with a focus particularly on assessment and competence are discussed, drawing upon existing literature. The context of this study in Sub-Saharan Africa is explained in Chapter 3, and Chapter 4 comprises a detailed account of all aspects of the research methodology for this study. The rich findings from all three study sites are brought together and presented in Chapter 5. The interpretation and synthesis of these findings follow on in Chapter 6, constantly referring back to the aim of the study concerning the validity of exit-level assessment

practices in determining the clinical competence of medical graduates in a resource-constrained context such as Sub-Saharan Africa. This final chapter concludes with limitations, recommendations for practice and future research, and my personal reflections.

Chapter 2

Theoretical Perspectives

Overview of Chapter

In this chapter, the subject of assessment is explored primarily from the medical education perspective. This starts with the definition of assessment and briefly, its relation to curriculum, and then proceeds to look at some purposes of assessment. Assessment at exit-level in medical programmes is usually summative, and the focus is on the clinical competence of graduates. In light of the contemporary focus on competency-based education, the difference between "competence" and "competency" is explored briefly. Returning to clinical competence, methods for its assessment are reviewed. It is important to ensure the quality of assessment, especially at exit-level from a medical-training programme, as this has implications in the licensing and certification of graduating doctors. Criteria for good assessment will therefore be discussed next. However, in settings where there are severe constraints on available resources, applying these criteria is not straightforward and can be influenced by available resources.

2.1 Understanding assessment

Although assessment is an integral part of the educational process in academic settings the term "assessment" has been described as "a problematic word" (Knight, 2006:436). A search for the meaning of "assessment" in dictionaries provided descriptions that involve some process of measurement and making a judgement (Cambridge University Press, 2013; Chambers, 2014; Harper Collins, 2014; Merriam Webster, 2015). In medical education circles, the subject of assessment is an important one (Rotgans, 2012), and has been well-researched as indicated by the volume of literature on the topic. However, although there are a number of definitions of this term these do not necessarily imply agreement. Table 2.1 summarises some descriptions found in the medical education literature.

Table 2.1 Descriptions of the term "assessment" in medical education literature

Source	Description of "assessment"
Wojtczak (2002:216)	"A system of evaluation of professional accomplishments using defined criteria and usually including an attempt at measurement either by grading on a rough scale or by assigning numerical value. The purpose of assessment in an educational context is to make a judgment about the level of skills or knowledge, to measure improvement over time, to evaluate strengths and weaknesses, to rank students for selection or exclusion, or to motivate."
General Medical Council (2009:94)	"All activity aimed at judging students' attainment of curriculum outcomes, whether for summative purposes (determining progress) or formative purposes (giving feedback). An 'examination' is an individual assessment test."
Downing and Yudkowsky (2009:11)	"Assessment" very broadly includes "any method, process or procedure used to collect any type of information or data about people, objects or programs" [From: (American Educational Research Association et al., 1999)]
Norcini et al. (2011:206)	"Assessment involves testing, measuring, collecting, and combining information, and providing feedback."
Pangaro and ten Cate (2013:e1209)	"Assessment" refers to "the process of making observations about the learner's proficiency, and comparing these to a standard" "Evaluation" refers to "the process of making a judgment that gives meaning to observations about the learner's proficiency, usually by comparing expectations"
Friedman Ben-David (2013:303)	"Assessment" in terms of "Formative assessment" (assessment "in order to intervene with intent to improve") and "Summative assessment" (assessment "in order to make such decisions as good/bad, ready to move forward, repeat a programme").
Schuwirth and van der Vleuten (2010:243)	"…any *purported* and *formal* action to obtain information about the competence and performance of a candidate…"

These descriptions have points in common and suggest there is a process involved in assessment, comprising several steps which include making observations and measuring to collect information, making comparisons with pre-determined criteria or standards, followed by forming a judgement summatively. Put in another way by Gitomer and Zisk (2015:3), "Assessments can be thought of as having three components – a model of cognition (what we are trying to measure), a model of observation (how we collect evidence) and a model of interpretation (how we make sense of the evidence)".

There is also some purpose to assessment, such as providing feedback or to make a decision on the student's performance. This will be discussed further in the following Section 2.2.

A word of caution is required here about the terms "assessment" and "evaluation", which both occur in the literature. In common parlance, "assessment" can appear to be synonymous with "evaluation". However, in an educational context, they may have very different meanings, depending on the accepted usage in a particular country (Taras, 2005). For the purposes of this thesis, "assessment" refers to all activities designed to gauge learner achievement or performance following an instructional programme, and "evaluation" refers to activities designed to gauge the effectiveness of the instructional programme or delivery of the course. This choice of usage has also been made taking into consideration the way in which these terms are used in the countries that were selected for this study.

In understanding assessment, it also needs to be considered with regard to the curriculum (curriculum being the overall content taught in an educational programme or course), since what is assessed is based on the education or training that has been provided in a curriculum within the programme of study. So the relation of assessment to the curriculum will be briefly described next, before proceeding on to a discussion on the purpose of assessment.

2.1.1 Assessment in relation to curriculum

In Chapter 3, it will become apparent that the education of health-care professionals needs to be aligned to the health-care needs of the local community. In order to provide education that has relevance in application, the curriculum and assessment also need to be aligned. In describing constructive alignment, Biggs (1996) referred to it as an approach to teaching in which the intended learning outcomes in an educational programme are first defined and used to align teaching strategies and assessment (Biggs, 1996; Biggs, 2014). Although the concept of constructive alignment has existed in the literature for some time, it is often associated with the work of Biggs.

2.2 Purposes of assessment

Assessment needs to be undertaken with a specific objective or purpose in mind, as this could influence the choice of assessment method(s) (Crossley et al., 2002; Epstein, 2007). From a higher education perspective, Natriello (1987:157) identified "four generic functions" when considering purpose, namely, certification, selection, direction and motivation, whilst in Newton's opinion, there were at least eighteen categories of purpose for assessment (Newton, 2007:161-162), depending on the user and "the decision, action or process supported by the assessment results". Following on from this perspective, in medical education there has been the use of the terms "formative" and "summative" when considering assessment (Wass et al., 2001c; Epstein and Hundert, 2002; Epstein, 2007; Holmboe et al., 2010). In addition to assessment directed at students, assessment has also been used to provide information to teachers and institutions on the quality and effectiveness of teaching (Wojtczak, 2002; Norcini et al., 2011). There are parallels between the descriptions of the purposes of assessment as described by Natriello (1987) in higher education and those that occur in medical education as described above and to follow, and these are summarised in Table 2.2 to illustrate some of these parallels.

With reference to Table 2.2, Natriello's four "generic functions" or purposes of assessment are displayed in the left half of the table, together with his description of each function. On the right half of the table, similarities of purposes of assessment from various medical education sources have been listed corresponding to Natriello's "generic functions". For example, "summative assessment" and the "assessment of learning" speak to Natriello's "certification", whereas "formative assessment" and the "assessment of learning" correspond to his "direction". Providing information to teachers and institutions on the quality and effectiveness of teaching enables them to engage in diagnosis and further planning, which is also part of Natriello's "direction". "Selection", concerning the act of progression along an educational path with satisfactory progress, can also be considered to have a summative purpose (As an example, deciding whether or not a student progresses to the next year of study or not). This brief outline here is simply to highlight several parallels in purposes of assessment as described in some higher and medical education literature. The focus now needs to return to the medical education point of view.

Table 2.2 Comparison of some descriptions of the purposes of assessment as described by Natriello (1987) in higher education literature and various sources in medical education literature

Higher education literature: Natriello (1987)		Medical education literature: Various sources	
"Generic function"	**Natriello's description**	**Descriptions**	**Sources**
Certification	Assurance that a student has attained a certain level of accomplishment or mastery	• Summative • Assessment of learning	Wass et al. (2001c) Epstein and Hundert (2002) Shumway and Harden (2003) Epstein (2007) Holmboe et al. (2010) Schuwirth and Van der Vleuten (2011) Dannefer (2013)
Selection	Identification of students or groups of students to be recommended or permitted to enter or continue along certain education and occupational paths	• Progression to the next stage of the course or training programme if performance has been satisfactory • Summative	Barzansky and Etzel (2003) Howe et al. (2004) Hauer et al. (2005) *[As listed above]*
Direction	Use of evaluation processes to communicate the specific desires of evaluators to those being evaluated and to allow evaluators to engage in diagnosis and further planning	• Formative • Providing information to teachers and institutions on the quality and effectiveness of teaching • Assessment for learning	Wass et al. (2001c) Epstein and Hundert (2002) Epstein (2007) Holmboe et al. (2010) Wojtczak (2002) Norcini et al. (2011) Schuwirth and Van der Vleuten (2011) Dannefer (2013)
Motivation	Engaging those being evaluated in the tasks in hand		

The terms "formative" and "summative" were first used by Scriven (1967) with regard to programme evaluation and improvement of the curriculum, but Bloom (1971) extended the usage to the process of student assessment, which is the context for its generally accepted current meaning (Newton, 2007). Although there is substantial literature in higher education on the application of these two terms (Sadler, 1989; Wiliam and Black, 1996; Boud, 2000; Taras, 2005; Boud and Falchikov, 2006), their usage from a medical education perspective will be discussed here.

Providing feedback to the student is a key task with formative assessment (Wojtczak, 2002; Epstein, 2007; Norcini et al., 2011). Summative assessment, on the other hand, usually occurs at the end of the course or training programme, and provides information on how much the student has learned (Wojtczak, 2002:216) or as a measure of the student's "educational achievement" (van der Vleuten, 1996). In more recent times, there has been a move to describing formative assessment as "assessment *for* learning" (Schuwirth and Van der Vleuten, 2011; Dannefer, 2013). Correspondingly summative assessment is depicted as "assessment *of* learning".

In medical education, the use of summative assessment for certification at the end of a programme is a crucial function of assessment. As this study focusses on summative assessment at exit-level, this will be discussed further in the next section.

2.3 Summative assessment for certification in medical education

The basic medical programme for doctors differs worldwide, with variables including duration and entry level. There is the more traditional entry of secondary (or high) school leavers to an "undergraduate" five- to six-year programme, or the shorter four-year "graduate-level" programme for individuals who already have a first degree (Searle, 2004; Sefton, 2004). The terminology used to describe the medical degree awarded also varies considerably and can be confusing but usually follows historical precedents. As an example, the Bachelor of Medicine, Bachelor of Surgery (abbreviated as MBBS) professional degrees are both awarded upon graduation from the relevant medical school by institutions in the United Kingdom. The Latin equivalent of the MBBS degree would be *Medicinae Baccalaureus, Baccalaureus Chirurgiae* (abbreviated as MBChB) and in addition to being used by some British medical schools, it is also common in South Africa. Countries which have followed British conventions with

regard to the style of naming of professional degrees would probably include former British colonies as well as other countries in which there has been a strong British influence in the past. In North American and European countries the equivalent medical degree awarded is Doctor of Medicine (abbreviated as MD). In this study, all of the medical schools studied have undergraduate first-degree medical programmes with students entering the programme from secondary school. This brief explanation regarding the style of naming the medical degree is simply to provide some background, and will have relevance later when the participating medical schools in this study are introduced in Section 4.3.1 and their respective medical degrees are illustrated in Table 4.1.

Summative assessment may take place at periodic intervals throughout the programme to measure the student's progress through the curriculum (Shumway and Harden, 2003). Typically, the student can only progress to the next stage of the course or training programme if their performance has been satisfactory (Barzansky and Etzel, 2003; Howe et al., 2004; Hauer et al., 2005).

The summative assessment which takes place at the end of a course or training programme, which could also be termed "exit-level assessment", certifies the student's level of attainment (Boud and Falchikov, 2006) with a tone of finality. As exit-level assessments are used for licensing and certification purposes (Epstein, 2007), they are regarded as "high-stakes" in nature (Knight, 2002), with significant implications for the student, the curriculum, the institution and the public (Newble, 1992; Wass et al., 2001c; Epstein and Hundert, 2002; Roberts et al., 2006; Epstein, 2007; Norcini et al., 2013). Medical schools have a responsibility to ensure and demonstrate that a certain level of competence has been achieved in their graduating doctors (Shumway and Harden, 2003), as this is central to public accountability. The protection of patients is of paramount importance (Norcini et al., 2011), and in addition, given the recent trends in medical litigation due to malpractice rising to worrying levels (Danzon, 1990). Medical schools are also increasingly being challenged to provide evidence that the assessments used are discriminating between the sufficiently and minimally competent at all levels of medical training (General Medical Council, 2009; 2011; Liaison Committee on Medical Education, 2013). This issue of "minimally competent" or "borderline" was highlighted in Boursicot and Roberts' study (2006). In such high-stakes situations, the

assessment process has to be rigorous and consistent with reference to standards that have been set by national licensing bodies, so that credible and defensible pass-fail decisions can be made (Norcini, 2003b) for students and for the health system into which they will move.

2.4 Summative assessment of clinical competence

2.4.1 Understanding clinical competence

As mentioned earlier, the assessment of clinical competence in the medical programme is a key task for medical teachers to certify a level of achievement (Newble, 1992; Wass et al., 2001c; Epstein, 2007), especially at the end of the programme. Before exploring clinical competence further, some discussion about the term "competence" is necessary in order to make sense of the existing literature on this subject. The dictionary meaning of "competence" refers to the ability or capability to do something well (Chambers, 2014; Harper Collins, 2014; Oxford University Press, 2015). In these sources of information, "competency" is deemed to be synonymous with "competence".

In the literature, the definition varies depending on the context and the term is used in fields as diverse as education, linguistics, biology, psychology, management, human resources, and politics (Taylor, 1988; Rowe, 1995; Hoffmann, 1999; Moore et al., 2002; Le Deist and Winterton, 2005; Jeris et al., 2005; Fouad et al., 2009). In a management or organisational context, "competence" is often taken to refer to skills and the standard of performance, whereas "a competency" refers to the behaviour by which it is achieved (Rowe, 1995; Moore et al., 2002). Rowe (1995) suggests a simple way of distinguishing between these two terms: "competence" describes what people can do, and "competency" relates to how they do it.

A substantial amount of research has been done in the field of nursing with respect to the assessment of competence (Watson et al., 2002; McMullan et al., 2003; Cowan et al., 2005; Yanhua and Watson, 2011). McMullan et al. noted that "terms such as competence, competency, capability and performance [were] being used inconsistently and often interchangeably" (McMullan et al., 2003:285). Prior to the last decade, there did not appear to be much consensus about the definition of "competence" with regard to nursing practice (Watson et al., 2002; Cowan et al., 2005). However, since then, Yanhua and Watson claim that there has been progress towards consensus about the

concept of clinical competence in nursing, although their paper does not provide any updated definition of "competence" (Yanhua and Watson, 2011:832).

From the medical perspective, various attempts have also been made to provide a definition (Epstein and Hundert, 2002; Carraccio et al., 2002; Govaerts, 2008; Albanese et al., 2008; Frank et al., 2010a; Fernandez et al., 2012). However, there is likewise a lack of consensus over what the term means. Epstein and Hundert's definition of "professional competence" as being "the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served" (Epstein and Hundert, 2002:226) would therefore be in agreement with the concept of combining knowledge, skills and "other components" (as defined in the following paragraph). Conversely, Caraccio et al. (2002) and Albanese et al. (2008) supported the view of selecting or specifying measureable behaviours for achieving competence.

Fernandez et al. (2012) conducted a systematic review of literature to analyse how health-science educators define competence, and argued that there are essentially two diverging conceptions on competence. These could be described in terms of their components, the use of the components and also the assessment of the components. There was agreement that competence was composed of knowledge, skills and "other components" (Fernandez et al., 2012:360). These "other components" were to do with abilities, such as reflection and "professional socialisation" (Fernandez et al., 2012:360), or attitudes, such as "professional judgement", and "character attributes" (Fernandez et al., 2012:360), which were based on values and judgement. Although there was no agreement on which "other components" were essential, values and attitudes were deemed a necessary part of competence. In terms of the use of the components, one concept related to the components of competence being combined "as a function of a complex situation" (Fernandez et al., 2012:361) in order for the individual to be competent, whereas the other concept required an appropriate selection from the available components for the complex situation.

Moving on to assessment of the components of competence, Fernandez et al.'s (2012) analysis found that for competence as a combination of components, the focus of assessment was on the way the student assimilated what had been taught (and what they had learned) and applied it to the clinical context that they were dealing with in

order to make "professionally meaningful actions" (Fernandez et al., 2012:363; Hawkins et al., 2015). This concept would sound familiar in the more "traditional" and time-based medical programmes, where there is attention to the process of education (in terms of the curriculum, teaching and learning) (Hodges, 2010). However, recent shifts in medical education have seen a focus on competence or competencies (Frank et al., 2010b) with "competency-based education" (as it has become known), the focus is on the end product of the training – the graduate – and what (s)he is capable of doing.

Competency-based education started in the 1950s drawing on behaviourist roots when educational theory was strongly influenced by psychologists (Brooks, 2009; Morcke et al., 2013). Emphasis is placed on educational outcomes. However, it has, so far, remained unclear how learning outcomes have impacted on teaching and learning, and whether competency-based education has resulted in more competent doctors (Morcke et al., 2013). There are concerns and reservations about competency-based education in terms of its educational approach and that it is not applicable to the training of the medical profession (Grant, 1999; Leung, 2002; Talbot, 2004; Rees, 2004; Norman, 2006; Hodges, 2006; Huddle and Heudebert, 2007; Brooks, 2009; Hodges, 2010; Malone and Supri, 2012; Brightwell and Grant, 2013; Morcke et al., 2013).

Critics have argued that competency-based education came about primarily for political and societal reasons because of pressure from outside the profession to demonstrate visible public accountability (Grant, 1999; Brightwell and Grant, 2013). It was initially introduced for the training of manual trades, so the focus of assessment was mainly on knowledge and observable skills (Malone and Supri, 2012) that were translated into competencies. In some settings, individual skills or attributes can be measured in isolation, and perhaps more easily "as a series of discrete task descriptions". However, attitudes and values, or the "other components" as designated by Fernandez et al. (2012), were not observable, and therefore were not measured or assessed. Huddle and Heudebert referred to "bite-sized elements of performance……identified by discrete learning objectives" (Huddle and Heudebert, 2007:538) which were assessed in a fragmented way and that this did not resemble what students would encounter in real clinical situations. Other authors have argued that the holistic practice of medicine is more complex than just being able to perform discretely defined tasks to a certain level of proficiency adequately, and that this professional practice encompasses higher order

cognitive and performance skills which do not fit into competencies (Grant, 1999; Malone and Supri, 2012; Brightwell and Grant, 2013).

Competency-based medical education has been advocated in many medical schools across the world, notably in the United Kingdom, Western Europe, United States and Canada, Australia and New Zealand. National associations and regulatory bodies for the medical profession have provided frameworks with learning outcomes and competencies to guide medical schools (Anderson et al., 1999; Royal College of Physicians and Surgeons of Canada Working Group, 2000; Simpson et al., 2002; General Medical Council, 2009; Australian Medical Council, 2012; Health Professions Council of South Africa, 2014).

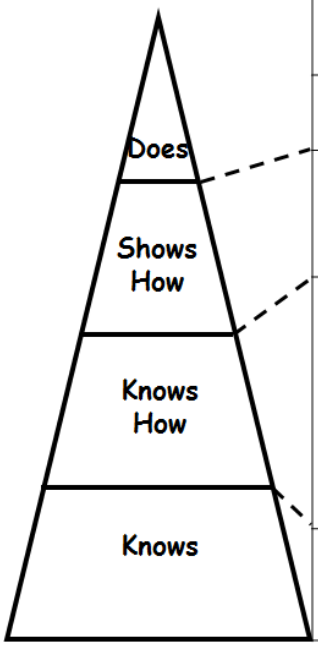
In some circles, competency-based education has been regarded as synonymous with outcomes-based education (Morcke et al., 2013) whereas others have contended there is a subtle difference depending on the perspective of the teacher or the student (Cumming and Ross, 2007; Albanese et al., 2008). Albanese et al.'s view is that an outcome identifies and defines "the skills and qualities we want doctors to have" whereas a competency "starts with a focus on patient care outcomes and takes the additional step of determining which outcomes doctors need to have" (Albanese et al., 2008:251). Outcomes-based education is also mentioned here since this is the education approach introduced as curriculum reform into the South African context following the country's first national democratic elections in 1994 (Department of Education (DOE), 1997; Jansen, 1998; Cross et al., 2002). While this is not the focus of this study, it does draw attention to the relevance of learning outcomes and competencies in a Sub-Saharan African context and calls into question gaps that exist between high ideals at policy level and what is achievable realistically in this given setting.

Taking into account all the intricacy of contemporary arguments about competence and competency, Miller (1990) proposed a simple approach to competence and the assessment of competence. It has come to be known as Miller's 'pyramid' and is widely espoused as a guiding framework in the assessment of clinical competence (Wass et al., 2001c; Crossley et al., 2002; Schuwirth and Van der Vleuten, 2003b; Wilkinson and Frampton, 2004; van der Vleuten et al., 2010). George E. Miller (1919-1998) was a pioneer of reform in medical education. His original paper "The assessment of clinical

skills/competence/performance" (Miller, 1990) was written as an invited review article providing a perspective on medical education at the time. Based on astute observations that he made over the course of his professional career, Miller used a two-dimensional pyramidal structure divided into four horizontal tiers to illustrate the complexity of assessing clinical competence, moving upwards from reproduction or factual recall of knowledge in the lower tiers to demonstration and application at the summit. This structure, termed Miller's 'pyramid', provides a reference framework that has come to be widely accepted and cited (See Figure 2.1). It starts at the base tier with knowledge or what the student "knows" (which could be assessed, for example, with oral examinations or basic written tests such as multiple choice questions or essays). The next tier up relates to "knows how" or the application of knowledge (assessed, for example, by using problem solving questions or short-answer type questions). The following tier, "shows how" requires the student to demonstrate that they know and know how, but can also show how to perform a given task or skill (for example, in an objective structured clinical examination, with the use of simulated patients, or in long and short clinical cases). The final tier at the summit of the pyramid relates to what the student or graduate "does" in actual clinical practice. Miller's 'pyramid' could serve as a useful framework to guide the choice of assessment methods in relation to the stage of training. For example, in the earlier foundation years of medical training, as the student is building up a broad knowledge base, assessment methods focus on the factual recall of knowledge and its application in a clinical context (Wass et al., 2001c). Later as the student gets increasing exposure to clinical settings, adding to the knowledge base and learning essential skills, assessment methods should move progressively up the levels of the pyramid. By the end of the medical programme, summative assessment should therefore be focussing more on the "shows how" and "does" levels.

Framework for clinical assessment

The Assessment of Clinical Skills/Competence/Performance



Tier	Miller's description (of requirements for the delivery of professional services by a successful physician)	Measurement	Assessment methods
Does	Functioning independently in a clinical practice	Action	Direct observation in a clinical setting
Shows How	Knows and Knows How, but can also show how to do it	Performance	Clinical simulations Use of patient substitutes (SPs) OSCE
Knows How	Use the knowledge accumulated to develop the skills of acquiring information from a variety of human and laboratory sources to analyse and interpret those data → translate findings with rational diagnostic or management plan	Competence	MEQs PMPs Computerised simulations
Knows	Know what is required in order to carry out professional functions effectively	Knowledge	Objective test methods

Figure 2.1 Framework for Clinical Assessment according to Miller (1990)

Notes:

(1) MEQs and PMPs are two assessment methods that have now fallen into disuse although they would have been acceptable at the time Miller wrote this paper. MEQs are difficult to construct well, and although they were introduced to test problem-solving skills, they tended to test knowledge recall only (Palmer and Devitt, 2007; Palmer et al., 2010). PMPs and MCQs measure the same thing, and MCQs appear to be more reliable, valid and efficient than PMPs (Norcini et al., 1985).
They have both since been replaced by well-constructed MCQs (Norcini et al., 1985; Palmer and Devitt, 2007; Palmer et al., 2010).

(2) In Miller's paper, the original pyramid structure showed each tier prominently labelled "Knows", "Knows How", "Shows How" and "Does", with the respective measurements "Knowledge", "Competence", "Performance" and "Action" added in small letters in parenthesis below. Miller used the description "quality of being functionally adequate, or of having sufficient knowledge, judgment, skill, or strength for a particular duty" to explain "competence". Although this does not appear to concur with the preceding discussion of competence here in Section 2.4, it is left as is in Figure 2.1 to show its original position in his pyramid. Reference to Miller's 'pyramid' nowadays focuses on "Knows", "Knows How", "Shows How" and "Does" and not the words he wrote in parenthesis.

Abbreviations:

OSCE	Objective Structured Clinical Examination
SPs	Standardised patients
MEQs	Modified Essay Questions
PMPs	Patient Management Problems

Figure adapted from Miller (1990)

So at this point, although it has been difficult to define competency, and the debate continues for and against competency-based education, it might be worth focussing on the principle that the ultimate goal of the medical-training programme is to have a doctor who is able to practise medicine in a manner that is "for the benefit of the individual and community being served" (Epstein and Hundert, 2002:226). The perspective I use of competence for this study would follow Epstein and Hundert's definition, as my ultimate concern in determining the clinical competence of a student is whether (s)he will be able to practise satisfactorily as a doctor in his/her local setting.

2.4.2 Assessment of clinical competence

At the final-year exit-level of undergraduate medical programmes the assessment taking place should be the assessment of clinical competence. In terms of how clinical competence is assessed, an extensive range of methods is available, as reflected in the literature (Wass et al., 2001b; Barzansky and Etzel, 2003; Shumway and Harden, 2003; Howe et al., 2004; Wilkinson and Frampton, 2004; Roberts et al., 2006; Epstein, 2007; Norcini and McKinley, 2007; Kogan et al., 2009; van der Vleuten et al., 2010; Jelovsek et al., 2013). The methods can be used for formative as well as summative purposes of assessment. Many institutions typically use assessment methods in combination (Fowell et al., 2000; Mavis et al., 2001; McCrorie and Boursicot, 2009; Ingham, 2011).

2.5 Assessment methods commonly used for exit-level summative assessment

Thus far, assessment has been discussed generically, leading to the medical education perspective, and including the assessment of clinical competence. This study, however, is about exit-level assessment and an understanding of the strengths and weaknesses (or limitations) of individual methods will provide a useful reference base with which to analyse collected data later. The strengths and limitations take into account criteria for quality in assessment (which will be discussed further in Section 2.6). A selection of different methods of summative assessment commonly used at exit-level are discussed here to provide a sense of the range that is available, and the relevance of this discussion will become apparent later with reference to the actual assessment methods that were encountered in this study. Table 2.3 at the end of this section summarises the main points in the ensuing discussion of assessment methods and also provides a

correlation of the method described with the corresponding tier of Miller's 'pyramid'. This correlation will also have relevance later in this study.

Written assessment formats such as essays and short answer questions are still used in most medical schools (McCrorie and Boursicot, 2009), although they are time consuming to mark and require much examiner time. This may be challenging as practising clinicians make up a sizeable proportion of faculty in many medical schools, and if clinical service work places heavy demands on their time, the amount of remaining time devoted to teaching and assessment may be reduced. **Essays** have been a traditional method, often requiring students to write in full sentences, rather than the "point form" that is more acceptable with short answer questions. The question then arises whether essays would be testing the student's abilities at written communication rather than aspects of clinical competence. Marking can also be inconsistent, with possible examiner subjective bias (for example, if the examiner was the subject specialist for the topic of the question being marked). It is argued that essays can only test limited-content area, leading to poor reliability (Wass et al., 2001c; Schuwirth and van der Vleuten, 2003a; Schuwirth and van der Vleuten, 2004). **Short-answer questions (SAQs)**, on the other hand, have been described as being able to test more content area than essays, thereby improving its reliability, and perhaps also being quicker and easier to mark, making it more feasible. However, typically SAQs still take more time to answer than multiple-choice questions, so there is less opportunity to sample broadly, and they are less suitable for assessing factual knowledge (Schuwirth and van der Vleuten, 2003a). In terms of feasibility or practicality (which will be discussed further in Section 2.6), these two methods of written assessment are generally fairly inexpensive, with costs including stationery and the printing of examination papers, but the availability of a suitable venue in which to conduct the examination for large numbers of students needs to be taken into account. In terms of issues of manpower and time, consideration also needs to be given for examiner availability and marking time as well as administrator time to capture the individual marks.

Multiple-choice questions (MCQs) and other written assessment formats, such as **extended-matching questions (EMQs)** and **key-feature questions (KFQs)**, have the potential advantage of also being conducted in a computerised format, if the medical

school has these resources available, or in the old fashioned "paper and pencil" format. Marking of examination papers can also be computerised, reducing the time taken and demand on examiner as well as administrative manpower. However, there are a number of challenges to consider. Multiple True/False type MCQs are easier to write than 'Single-best answer' type MCQs, but can be more problematic (Case and Swanson, 2002). A question (or item) that is apparent to the item writer at the time of writing the question may not be so on review, and ambiguities may appear when interpretation varies. When attempts are made to reduce the ambiguities, the process may result in items that just assess the recall of isolated facts. However, this does not mean they are not still in use in medical schools round the world. Single-best answer MCQs are seen as being better for the application of knowledge, integration, synthesis and judgment questions (Case and Swanson, 2002), with a move away from the multiple true/false type MCQs (Anderson, 2004; McCoubrie, 2004).

Extended-matching questions (EMQs) and key-feature questions (KFQs) have been in use only more recently, and are examples of "context-rich" question types (Schuwirth and van der Vleuten, 2004), as are single-best answer-type MCQs. "Context-rich" refers to the case scenario or vignette that accompanies the question. These case scenarios are often written based on true cases or incidents and can help to provide a more realistic setting with which to test the student's application of knowledge and problem-solving skills. The questions are time consuming to construct, and question writers need to be trained first for any form of multiple-choice question, although EMQs are easier to construct than KFQs. Both are good for testing the application of knowledge and problem solving.

The **long case** has long been used as a popular method of assessment for many years, as it provides an opportunity to assess the student's ability to gather information and develop a plan of management in a more authentic context (Wass and van der Vleuten, 2004; Ponnamperuma et al., 2009). Usually, one case (a real patient) is allocated to one student. This however, may give rise to concerns about case specificity. It has long been recognised that the performance of doctors in one context does not predict performance in other contexts (Swanson et al., 1995; van der Vleuten, 2000; Norcini, 2002). There is also the issue of feasibility, as it means that a large number of suitable cases would need to be found for a large class of students. It could also mean that the

degree of difficulty of the cases could vary considerably, leading to issues of reproducibility, examiner subjectivity and inter-case reliability (Gleeson, 1997; Norcini, 2002; Wass and van der Vleuten, 2004). There is a danger that long cases could easily turn into oral assessments unless there is prior examiner agreement on the purpose and conduct of the assessment and what aspects of clinical competence are being measured. The OSLER (Objective Structured Long Examination Record), a 10-item analytical record of the traditional long case, was developed in the 1990s in an attempt to improve the objectivity, reliability and validity of existing practices (Gleeson, 1997). It provides a more structured presentation of an unobserved long case to guide examiners on what needs to be measured, and includes some direct observation of the student interacting with the patient. Other ways of increasing the reliability of long cases that have been tried have included increasing the number of cases used per student, increasing testing time, and using simulated patients instead of real ones (Wass et al., 2001a; Norcini, 2002; Wass and van der Vleuten, 2004). The implications here would include issues of the costs incurred with such exercises, the manpower involved, and also in terms of training of examiners as well as patients (for consistency). These challenges to meet these issues would be even greater where resources are scarce.

The **OSCE (Objective Structured Clinical Examination)** was introduced in the 1970s as a more standardised way of assessing clinical competencies, especially at the "Shows how" level of Miller's pyramid (Harden et al., 1975; Harden and Gleeson, 1979). Typically candidates are assessed at a number of stations on activities that simulate aspects of competence, such as history taking, physical examination, communication skills with patients and family members, and the demonstration of psychomotor skills (Shumway and Harden, 2003). The choice or selection of the focus of each station could be determined by a process of blueprinting (which will be discussed further in Section 2.6). The candidates rotate through each station in sequence in the same specified time. Clinical models and simulated patients or standardised patients can be used to allow large numbers of students to be tested on the same clinical problem without causing fatigue or stress to real patients. Direct or indirect observations as well as checklists and rating scales measure the performance against predetermined standards resulting in a more objective examination than with traditional methods.

Over the past 40 years since its introduction, OSCEs have become widely popular and have been well-researched (The small selection of references listed here provide some indication of the many perspectives relating to OSCEs)(Harden and Gleeson, 1979; Wass et al., 2001b; Hodges et al., 2002; Davis, 2003; Roberts et al., 2003; Wilkinson et al., 2003; Humphrey-Murto et al., 2005; Jefferies et al., 2007; Pell et al., 2010; Johnston et al., 2013; Patrício et al., 2013; Sandilands et al., 2014). The simulated format of the assessment with the use of structured checklists or rating scales has helped to improve reliability not found in other unstructured performance-based assessments such as the long case. Validity would depend on factors including the number of stations, the selection of cases or scenarios for the stations to ensure broad sampling and the cut-off score. Although highly feasible, OSCEs can be expensive to administer (Patrício et al., 2013). Costs include the equipment and facilities required to conduct the assessment, time and effort to devise and write stations, checklists and/or rating scales, training of simulated patients, and examiner time. There is also the added dimension of coordination of OSCEs which would require a level of expertise and experience on the part of the individual who is coordinating (This task would usually be performed by a clinician). Such issues have implications especially where resources are limited, and this will be discussed further in Section 2.7.

The **OSPE (Objective Structured Practical Examination)** (Harden and Cairncross, 1980) is similar to an OSCE except it is used mainly for the assessment of practical skills in a non-clinical setting such as pathology and physiology. Although this would seem irrelevant in a study focussing on the clinical competence of medical students at exit-level, the use of "OSPEs" features later and an explanation of this assessment method is therefore provided here.

Clinical simulations reproduce clinical scenarios or situations in order to assess the competence of the student in those actual situations (Schuwirth and Van der Vleuten, 2003b; Norcini and McKinley, 2007), equating to the higher "shows how" and "does" levels in Miller's pyramid. The intention of the simulation is to provide as authentic an environment as possible in relation to what occurs in real clinical situations, thus enhancing the potential to make judgements about the student's competence to meet local health-care needs. Simulations could be computer-based, such as those used in computer-driven models or mannequins, and could also make use of simulated patients.

Both can be incorporated into an OSCE. Simulated patients (or SPs: Individuals who are trained to portray patients with particular medical conditions in an accurate and consistent fashion) can be used to test a student's ability to obtain relevant information from the patient (history-taking, communication and interpersonal skills) and physical examination, to summarise and interpret the information, and to plan the treatment and further management of the patient. Individual tasks that are tested are scored against pre-determined checklists or rating scales. However, large costs are involved in simulations (as already mentioned above with OSCEs). The reproduction of clinical details for authenticity adds to the expense, as does the use of computer-based mannequins and other computerised equipment and therefore has significant implications for resource-constrained contexts.

It could be argued that the **oral examination or viva** has been considered an integral form of assessment for the medical professions more for reasons of tradition than on educational grounds. It is said "to assess knowledge, to probe depth of knowledge and to test other qualities such as mental agility" (Davis and Karunathilake, 2005) and problem-solving skills. Although oral examinations have long been favoured as a method of assessment by examiners, they have very poor reliability mainly because of examiner bias which may include prompting. They are usually unstructured, which means there is inconsistency in what different students are tested on, and differences in the degree of difficulty of questions (Daelmans et al., 2001; Davis and Karunathilake, 2005) which can lead to concerns about fairness and reliability. Students can only be tested on a limited number of content areas in the time allocated for the oral examination, and the testing is confined mainly to the "knows" and "knows how" levels of Miller's 'pyramid'. This does not test the student's clinical skills and ability to "show how" or "do", which questions the validity of this form of assessment. Ways of improving the reliability of the oral examination include increasing the number of oral examinations (or testing time) and number of examiners, providing some standardised structure to the content areas that are tested and training the examiners (Davis and Karunathilake, 2005).

Other assessment methods have been developed for use in the workplace, such as the Mini Clinical Evaluation Exercise (Mini-CEX) and Direct Observation of Procedural Skills (DOPS). The feedback that can be given to learners is timely and specific, and this may

be valuable in changing their behaviour (Norcini and Burch, 2007). As with the aforementioned methods, such workplace-based assessment can be complex, relying on the development and use of checklists in some instances and judgements made on rating scales (that also need to be developed) in others (Crossley and Jolly, 2012). In the past they have tended to be used more in formative assessment situations and at postgraduate level, but are mentioned here as they are increasingly being introduced for summative assessment at undergraduate level and in internship (Hill et al., 2009; Weston and Smith, 2014).

Table 2.3 lists examples of assessment methods that have been divided into three main categories, written, clinical or performance-based assessments, and other assessments that do not fall under the previous two categories, with short descriptions provided for each method. This may be a simplified form of grouping and the list of methods is not exhaustive but illustrates summative assessment methods that are widely used. Each method has its strengths and limitations, and there may be overlap of some methods across the categories. The various assessment methods described are also assigned to the appropriate tier(s) of Miller's 'pyramid' to illustrate the correlation between these two elements, and reference will be made back to this link in Chapter 5.

Table 2.3 Examples of summative assessment methods and their correlation to Miller's 'pyramid'

Category	Assessment method	Short description	Correlation with Miller's 'pyramid'	Strengths	Limitations
Written	**Multiple-choice questions** (MCQs) [Ref: (Case and Swanson, 2002; Schuwirth and van der Vleuten, 2003a)]	The student is required to select an answer from a number of choices that are provided. There are generally two kinds of MCQs: True/False MCQs (that require the student to select all responses that are true) and Single-best answer MCQs (that require the student to select a single-correct response out of a series of responses).	KNOWS / KNOWS HOW	Both types of MCQs: • Suitable for assessing factual knowledge • MCQ test can cover a broad area • Concise • Can be answered quickly by the student One best answer MCQs: • Easier to construct than True/False type	True/False MCQs: • Difficult to construct without flaws • The student is required not only to make a judgement based on knowledge, but also guess the options that are not completely true or completely false One best answer MCQs: • Often used to test only facts
	Extended-matching questions (EMQs) [Ref: (Case and Swanson, 2002; Schuwirth and van der Vleuten, 2003a)]	These are multiple-choice items organised into sets that use one list of options for all items in the set. The EMQ has four essential components: a theme, an option list, a lead-in statement, and at least two item stems. Item stems provide clinical vignettes to which the student is required to select one best option from the list provided. There are also EMQs where students may be instructed to select a particular number of options, generally more than one.	KNOWS HOW	• Best used in situations where large numbers of similar sorts of decisions need testing for different situations (for example, relating to diagnosis or required laboratory investigations) • Easier to construct than key-feature questions	• Risk of an under-representation of certain themes if these do not fit the format

Category	Assessment method	Short description	Correlation with Miller's 'pyramid'	Strengths	Limitations
Written	**Key-feature questions** (KFQs) [Ref (Page et al., 1995; Schuwirth and van der Vleuten, 2003a)]	A description of a clinical scenario is provided followed by a number of questions that require essential decisions. The student is required to select from a list of options provided for each question. The "key feature" is the "critical step required in the resolution of a clinical problem".	KNOWS HOW	• Measures application of knowledge and problem-solving ability • Can be used also to test clinical decision-making skills • Valid • Has good reliability • Acceptable	• Not suitable for assessing factual knowledge • Time consuming to construct • Expensive to produce • Large numbers of cases are needed to prevent students from memorising cases
	Short-answer questions (SAQs) [Ref: (Webber, 1992; Schuwirth and van der Vleuten, 2003a)]	Open-ended questions, structured so that students construct their own concise answer (rather than choose preformed answers as in MCQs)	KNOWS / KNOWS HOW	• Tests knowledge and problem-solving ability	
	Essays [Ref: (Wass et al., 2001c; Schuwirth and van der Vleuten, 2003a)]	Open-ended questions in which the student can provide their answer usually without any limit on the length or number of words. Structured/unstructured	KNOWS / KNOWS HOW	• Tests how well students can summarise, hypothesise, find relations and apply known procedures to new situations. • Provide insight into different aspects of writing ability and the ability to process information	• Difficult to mark consistently • Time consuming to answer and to mark • High costs • Limited reliability

Category	Assessment method	Short description	Correlation with Miller's 'pyramid'	Strengths	Limitations
Clinical / Performance-based	**Long case** [Ref: (Wass and van der Vleuten, 2004; Ponnamperuma et al., 2009)]	A patient is selected (preferably with clinical signs although this is not an absolute requirement), from the ward or an ambulatory setting such as the outpatient department. The student is given approximately 30 to 60 minutes to interview the patient and perform a physical examination without the examiner(s) being present. The student then presents his/her findings to the examiner(s) and answers any oral questions around the findings, differential diagnosis, investigations and further management of the patient's condition.	KNOWS HOW / SHOWS HOW	• Assesses practical ability of student to gather and integrate relevant information from an interaction with a real patient (as in a real clinical encounter between doctor and patient) in order to arrive at a plan of management for the patient's condition(s) • Highly valid	• The patient is not trained and therefore there is no standardisation • Need a large pool of cases for a large class of students • Time consuming • High costs involved • Poor inter-case reliability • History-taking process is not observed
	Objective Structured Long Examination Record (OSLER) [Ref: (Gleeson, 1997)]	A student is given approximately 20 to 30 minutes with a selected patient (similar to the long case). This time is observed usually by more than one examiner, and assessed using a structured 10-item analytical record covering history taking, physical examination, investigation, management and clinical acumen. Each examiner assesses the student on the same 10 items.	KNOWS HOW / SHOWS HOW	• Attempts to standardise the long case • Can be used for assessing communication skills, especially in history-taking	• The patient is not trained • Need a large pool of cases for a large class of students • Time consuming • High costs involved

Category	Assessment method	Short description	Correlation with Miller's 'pyramid'	Strengths	Limitations
Clinical / Performance-based	**Objective Structured Clinical Examination** (OSCE) [Ref: (Harden and Gleeson, 1979; Hodges et al., 2002)]	Students rotate around a circuit of stations. At each station, there is a specific task to perform: usually a clinical skill in a simulated environment. A structured standardised marking scheme which has been determined in advance is used.	SHOWS HOW	• All students exposed to same clinical scenarios/presentations • Uses simulated patients trained to perform a task in a consistent way • Pre-determined standardised marking sheet or checklist used • Reliable • Fixed time for each stations	• Costly to conduct • Logistically challenging • Resource intensive – requires space to conduct, equipment, examiners, support staff, simulated patients and/or mannequins • Simulation of clinical scenario but the setting and timing may seem artificial
	Objective Structured Practical Examination (OSPE) [Ref: (Harden and Cairncross, 1980)]	Students rotate around a circuit of stations. This is an assessment of practical skills, knowledge and/or data interpretation in a non-clinical setting.	SHOWS HOW	• All students exposed to same clinical scenarios/presentations • Pre-determined standardised marking sheet or checklist used • Reliable • Fixed time for each stations	• Costs involved • Resource intensive – need time to set up, space to conduct, equipment
	Simulations [Ref: (Schuwirth and Van der Vleuten, 2003b; Norcini and McKinley, 2007)]	Simulation of clinical settings with simulated patients (SPs), the use of mannequins, or high technology (computer-based) can be incorporated into an OSCE and used for assessment.	SHOWS HOW / DOES	• More authentic (? or closer to authenticity)	• Expensive • Labour intensive • SPs need a high level of training

Category	Assessment method	Short description	Correlation with Miller's 'pyramid'	Strengths	Limitations
Clinical / Performance-based	**Mini Clinical Evaluation Exercise** (Mini-CEX) [Ref: (Norcini et al., 2003; Academy of Medical Royal Colleges, 2009)]	This takes place in a workplace setting and involves the direct observation of a student/patient interaction. The focus of the observation may be one part of the clinical encounter, such as an aspect of history taking or physical examination, clinical judgement and synthesis. The student is scored using a 9-point rating scale. Following the period of observation (usually approximately 15 to 25 minutes), there is an opportunity for feedback and discussion on the ratings with the student (approximately 5 minutes' time).	SHOWS HOW / DOES	• Assessment in the workplace • Can be used in formative assessment • Timely feedback to student	• Assessor (examiner) needs training on giving feedback
	Direct observation of procedural skills (DOPS) [Ref: (Academy of Medical Royal Colleges, 2009)]	The student is observed in a workplace setting performing a clinical procedure and scored using a 6-point rating scale, over approximately 15 minutes of observation time.	SHOWS /HOW / DOES	• Assessment in the workplace • Can be used in formative assessment • Timely feedback to student	• Assessor (examiner) needs training on giving feedback

Category	Assessment method	Short description	Correlation with Miller's 'pyramid'	Strengths	Limitations
Other	**Oral** [Ref: (Wass et al., 2001c; Davis and Karunathilake, 2005)]	A face-to-face interaction between the student and one or more examiners, usually without patients or clinical material.	KNOWS / KNOWS HOW		• Lack of standardisation of questions • Lack of testing time • Insufficient examiners • Examiner bias • Examiners require training • Low reliability
	Logbook [Ref: CanMEDS Assessment Tools Handbook]	Tool used to track incidence of educationally relevant activities e.g. number of procedures performed	DOES	• Assessment of activities in the workplace	
	Portfolio [Ref: (Davis et al., 2001; van Tartwijk and Driessen, 2009)]	A portfolio is a collection of papers and other forms of evidence that learning has taken place. It can be used for learning as well as assessment purposes. The material collected (in paper-based or electronic format) could include, for example, patient presentations, case discussions, practical procedures performed, learning plans, reports and assignments. Assessment is by a process of document review which may be undertaken together with an oral examination.	DOES		• Examiners need training • Time consuming

Category	Assessment method	Short description	Correlation with Miller's 'pyramid'	Strengths	Limitations
Other	**Peer assessment** [Ref: (Norcini, 2003b; Norcini and McKinley, 2007)]	An example is the mini-peer assessment tool (mPAT). The student nominates 8 assessors from among his/her peers and supervisors, which could include other health professionals. They each complete an assessment form concerning the student's technical and interpersonal skills.	DOES	• Assessment in the workplace	• Open to subjective bias of assessors

To summarise, in practice a range of assessment methods is used in the assessment of clinical competence because there is no single method that is able to test all aspects of the complex cognitive and performance skills that form the practice of medicine (Wass et al., 2001c; van der Vleuten et al., 2010). Each has its strengths and weaknesses, which need to be taken into account together with a number of other factors that influence the choice of assessment methods, including the purpose(s) of the assessment, the circumstances or context in which the assessment is conducted, and also a need to ensure quality in assessment. Criteria for sound assessment are used to help ensure the quality of individual assessment methods and could also be used to help decide the choice for exit-level assessment. This will be discussed in Section 2.6 to follow. The focus of this study is on exit-level summative assessment of the clinical competence of medical graduates in a resource constrained context such as Sub-Saharan Africa, and in Chapter 3, it will become apparent what impact resource constraints can have on health-care in general, and how this in turn will affect the teaching and learning activities taking place in medical schools in this region.

2.6 Criteria for quality in summative assessment

Graduating doctors need to achieve a level of competence that is crucial for certification. Quality assurance in summative assessment is important and necessary to ensure that certification decisions can be made with confidence and are defensible. This could also increase the confidence of stakeholders concerned, notably licensing bodies, the institution and the public. Safeguarding this quality in assessment needs to be considered from several perspectives. One perspective could be focussing on the quality of individual assessment methods. As mentioned before, a combination of a number of methods is often used for summative assessment, and a second perspective would be concerning the quality of the whole programme. These will be discussed in turn.

In order to make choices on the assessment methods that would best suit a particular exit-level assessment situation, there needs to be some way of ensuring the quality of that particular assessment method. There are criteria that are widely advocated, such as those proposed by van der Vleuten (1996) and Norcini et al. (2011). Van der Vleuten put forward a model to improve the usefulness of a particular method of assessment that has been highly cited over the years (van der Vleuten, 1996). This model took into

account issues of reliability, validity, impact on future learning and practice, acceptability to learners and faculty, and cost. More recently, drawing upon developments in assessment in medical education over the past 50 years or so, medical and other health-care professions educators have used conference gatherings to network and share ideas. The Ottawa Conferences in Assessment of Competence in Medicine and the Healthcare Professions, a biennial conference event that started in 1985, has been one such forum for discussion and debate on all aspects of the assessment of competence (Issenberg, 2011). The Ottawa 2010 Conference Planning Committee commissioned several consensus statements and recommendations drawing upon good practices in areas of assessment and identifying areas of priority and focus to inform future practices and further research. One of these consensus statements, from Norcini et al. (2011), provided a set of criteria for good or sound assessment which includes validity or coherence, reproducibility or consistency, equivalence, feasibility, educational effect, catalytic effect and acceptability. There is agreement on a number of the criteria from these two authoritative sources, and it could be argued they are both speaking to the same kinds of issues which arise when striving for quality in summative assessment and these therefore deserve further discussion. The two constructs of validity and reliability provide rigour to the assessment of student performance, and are emphasized especially in assessments conducted at exit-level (Wass et al., 2001c; Epstein, 2007; Wilkinson and Frampton, 2004).

Reliability concerns the reproducibility or consistency of a test or assessment in measuring learning consistently (Downing, 2004). Norcini et al. (2011) have taken this further with separate criteria of "reproducibility", referring to "the results of the assessment would be the same if repeated under similar circumstances" (Norcini et al., 2011:210) and "equivalence", referring to "the same assessment yields equivalent scores or decisions when administered across different institutions or cycles of testing" (Norcini et al., 2011:211).

The reliability of assessments of clinical competence needs to take into account variables that can affect it, such as the case(s) used in the assessment and the examiner(s) (Wass et al., 2001c). (The cases here refer to the use of clinical cases to illustrate a clinical condition or a clinical scenario. This could take the form of a written paper-based case or the use of simulated or real patients in a clinical examination.)

Inter-case reliability refers to the consistency of the student's performance across different cases. Making a judgement on the student's level of clinical competence would be more reliable if the assessment uses broad sampling across different clinical cases than just basing it on one case. Inter-rater reliability refers to the consistency of rating performance by different examiners, and this would need to be taken into consideration in situations where multiple examiners are used to assess the student's performance across different clinical cases. However, enhancing reliability has particular implications with regard to the availability of resources (including manpower, infrastructure, and accessibility of case material), posing challenges in settings which place more pressing demands on already stretched resources, and also questions the relevance in a given context.

Validity relates to the extent to which a test or assessment succeeds in measuring what it was designed to measure (Downing, 2003). Here in this study and within the context of Sub-Saharan Africa, it would relate to the validity of the decisions that are made about whether the students at the end of the medical programme have achieved the level of competence required for certification and to allow them to practise as doctors. These decisions would need to be based on the assessment practices that are undertaken and assessment content that is relevant for the context that the graduating doctors will practise in.

Categories of validity, such as content and construct, have been described in the literature (van der Vleuten, 2000; Shumway and Harden, 2003). For example, content validity refers to the alignment of sampled content with course content, with the testing of a good and adequate selection of topics. Construct validity refers to the ability of a measurement to differentiate between groups with known differences in ability, for example, between novices and experts.

A third criterion that also needs attention in view of the context of this study which takes place in a resource-constrained setting is feasibility. Norcini et al.'s (2011) criterion of **feasibility or practicality** of conducting good assessment needs to take into account available resources such as manpower, equipment and available space, but also investment in the training and standardisation of examiners, test-item construction, mechanisms to monitor the assessment processes and provide quality assurance. van der Vleuten's (1996)model refers to **cost**, and indeed, all these activities,

although highly desirable in the ideal world, come at a cost which may be ill-afforded where resources are limited. In practice, this ideal state is not always possible. One question that arises here, with reference to resource constrained settings, is whether the feasibility of conducting assessment in such a context would compromise the assessment practices that are undertaken, and this will be discussed further in Section 2.7.

Other criteria for sound assessment from van der Vleuten's and Norcini et al.'s models also have a role to play in ensuring the quality of individual assessment methods. These include acceptability and educational effect. However, as they are not the prime focus of this study they will not be discussed in further detail here.

The criteria discussed up to this point relate to assessment generically. Nevertheless, assessment at exit-level in medical programmes poses a particular set of challenges. The second perspective, mentioned earlier in this section, in ensuring quality in assessment concerned the quality assurance of the assessment programme, to provide a more comprehensive view of the assessment taking place. Measures that would contribute to the quality assurance include blueprinting and standard setting.

Blueprinting of an assessment can be used to map the assessment content in order to determine if intended learning objectives have been covered (Hamdy, 2006; Hays, 2008; Coderre et al., 2009). It can assist in ensuring content validity. Blueprinting would usually be conducted by the overall assessment coordinator and can provide a means of determining the breadth of coverage of the curriculum. As examples, assessment content can be mapped by organ systems or by clinical problems, or the test instrument used. Later in this study, a form of blueprinting that refers to Miller's 'pyramid' will be described in Chapter 4, and the findings obtained will be presented in Chapter 5.

Standard setting refers to a systematic process of "gathering value judgments, teaching consensus and expressing that consensus as a single score on a test" (Norcini, 2003b:464). The score in question, also referred to as a "cut point" or "cut score" is the passing score (Norcini, 2003b; McKinley and Norcini, 2014). However, there is no "best" or "right" passing score. Various methods are available for setting standards, and the choice of method depends on the purpose of the assessment. Where judgements have to be made on the competence or incompetence of a candidate, criterion-referenced or absolute standards are preferred (Norcini and Shea, 1997; Norcini,

2003b; De Champlain, 2004; Downing et al., 2006; Bandaranayake, 2008; McKinley and Norcini, 2014). Experts (or judges) are selected to make the judgements, and therefore it would appear there is a subjective element in their judgements. What is important is the number and choice of judges, and using a systematic approach to collecting their judgements (Norcini, 2003b; Downing et al., 2006). Although this process would contribute to ensuring the standards of assessments are defensible and maintained, it is evident the process of standard setting would probably have to compete for the very same meagre resources available in Sub-Saharan Africa in terms of human resources (such as examiners as expert judges, and psychometricians).

As a summary of this section, blueprinting and standard setting as well as criteria for sound assessment all contribute to ensuring the quality of assessment. However, in situations where resources are insufficient, applying these processes and criteria would be difficult, and it would not be unreasonable to surmise that assessment practices could be compromised. As resource constraints are very much in the foreground in a Sub-Saharan African setting, the following section will address further the implications for assessment.

2.7 Implications for assessment in a resource-constrained context

In the preceding discussion, reference has already been made to potential implications for assessment in a resource-constrained context such as Sub-Saharan Africa. In this section, the issue is explored in greater detail. Striving for good or sound assessment of clinical competence at the exit-level of a medical training programme is of great significance to medical schools because of the high-stakes nature of the assessment and the consequences on the licensing and certification of their graduates. The use of widely-accepted criteria (van der Vleuten, 1996; Norcini et al., 2011) should guide this process. However, when challenges are encountered such as those that arise in resource-constrained settings, how can we ensure that the quality of assessment practices is not compromised? Statistics from the 2006 WHO World Health Report indicated that Africa had just 1% of the total global funding for health while dealing with 24% of the global disease burden (World Health Organisation, 2006) with only 2.3 health-care workers per 1000 population compared to 24.8 per 1000 population in the Americas. As will be seen in Section 3.2, the human resources involved in assessment have high clinical workloads. So the demands made of faculty in terms of teaching as

well as clinical service are considerable. Medical schools in Sub-Saharan Africa face challenges with such acute shortages of resources, particularly when wanting to conduct resource demanding assessment methods such as the Objective Structured Clinical Examination (Reznick et al., 1993; Cusimano et al., 1994; Brown et al., 2015). Feasibility and cost of assessments may therefore have a considerable influence on the choice of methods and practices, which in turn, will impact on the validity and reliability of assessment.

In addition to human resources, the challenges of resource constraints in Sub-Saharan Africa also relate to poor infrastructure and shortages of equipment and medical supplies. While these issues genuinely exist and are based on data gathered and reported in studies, they are more often referred to in a collective way such as at the level of health policies rather than based on actual descriptions at source. Some corroborating evidence of these kinds of constraints have been found in qualitative studies, such as Frambach et al.'s study (2015) from two medical schools in Mozambique on the preparedness of students and junior doctors for clinical practice. Study participants who were interviewed described shortages of materials and medicine. They also mentioned "very rudimentary diagnostic tools" and how they had to rely on treating patients empirically without the means to conduct further investigations (or tests) to confirm the diagnosis (Frambach et al., 2015:70). Aveling et al.'s study (2015) was based in two East African hospitals in two different countries. While this study investigated issues of patient safety, deficiencies in the physical infrastructure as well as equipment and supplies were found to be one of the causes. Study participants referred to broken windows, malfunctioning doors, inappropriate building design, lack of beds, and unreliable electricity and water supplies, in addition to high patient volume (Aveling et al., 2015:3-4). Examples were also cited of demand exceeding supply, such as running out of sterilized packs because they were used up faster than anticipated. It is more than plausible that such working conditions would contribute to the challenges facing doctors and other health-care professionals who provide clinical services in such settings, and would consequently impact on the teaching and learning conducted by the same individuals. The challenges of human resources will be discussed further in Section 3.2.

Walubo et al. (2003) developed a hypothetical model to guide the selection of assessment methods for the Sub-Saharan African context. Four factors (ability to test for performance, cost, suitability and examination safety) that were identified as influencing the implementation of an assessment method were used in the evaluation of six selected assessment methods (essay, short-answer questions, multiple-choice questions, patient-based clinical examination, problem-based oral examination, and Objective Structured Clinical Examination). The factors were weighted according to importance and this was based on the setting of the medical school. Using their formula, the Objective Structured Clinical Examination (OSCE) was deemed the best assessment method for students' performance, but was most costly in terms of examiner manpower hours, equipment, venues and examination preparation. The converse held for written forms of assessment such as multiple-choice questions (MCQs) and short-answer questions (SAQs), although these may not be the optimal assessment methods. A problem-based oral examination fared better than the OSCE or a patient-based clinical examination, and essays and short-answer questions (SAQs) were equally more appropriate than multiple-choice questions (MCQs) as choice of assessment methods in an African medical school setting. Walubo et al.'s model does not appear to have been applied in any Sub-Saharan African medical school as a literature search has not yielded any published studies. More importantly, this study has highlighted the tension between having to make choices on assessment methods often on the basis of cost and availability of human resources, and the need to engage in sound assessment practices leading to certification of competence which requires resources that many Sub-Saharan African countries lack or can ill afford. This tension is one of the issues that this study seeks to explore.

As also mentioned earlier, faculty are involved in and responsible for the process and content of assessment. Given the resource constraints that currently exist in Sub-Saharan Africa, this would have implications on faculty development that would be essential and contribute to ensuring the quality of these aspects of assessment is maintained. The resource implications for assessment stretch beyond merely having staff available to construct and conduct the assessment. Many methods are technically demanding and require faculty development to ensure their appropriate use. This in itself has considerable resource implications. However, the indirect resource implications of faculty development are not the focus of this thesis.

There are also further implications on teaching and learning, since there should be alignment of assessment objectives with educational objectives in order to motivate the learning process (Biggs, 1996; van der Vleuten, 1996). Apart from the need to align assessment, particularly at exit-level, to the intended outcomes of the curriculum, this alignment should also include the health-care requirements of the respective country where graduating doctors will practise (Frenk et al., 2010). Given these different challenges and the complexity of the context within which assessment occurs, the question that arises is, whether exit-level assessments provide the information that is needed for judgements to be formed on the clinical competence of these medical graduates and their 'fit for purpose' to meet the health-care needs of the countries and regions in which they have trained.

There is a need to investigate what exit-level assessment practices are taking place in Sub-Saharan Africa, and how the choices are made, in order to determine the validity of the assessment for this resource-constrained context. This study has set out to explore this further. But first, some background on the lie of the land needs to be provided: some context on health-care and medical education in Sub-Saharan Africa will be presented in the following Chapter 3.

Chapter 3

Context of Study

Overview of Chapter

This chapter provides some geographical and historical context of the African continent and background for the study. It describes challenges of health-care in the Sub-Saharan African region. Resource constraints especially in terms of the shortage of health-care professionals are highlighted. With regard to measuring health-care, the use of the Millennium Development Goals and some epidemiological metrics from the Global Burden of Disease study are discussed. This provides a framework within which issues concerning the assessment of clinical competence in a resource-constrained context such as Sub-Saharan Africa can be further explored. It touches upon the origins and the gradual increase in numbers of medical schools in Sub-Saharan Africa and also the challenges of health-care in this region. This context, together with the discussion in Chapter 2, ultimately informed the development of a conceptual framework for this study. A preliminary study on the assessment of clinical competence in one Sub-Saharan African medical school is described, which involved the mapping of assessment practices taking place at exit-level. The findings of this pilot study then helped develop the research question for this study.

3.1 Defining Sub-Saharan Africa and Health-Care Issues

The continent of Africa covers an immense land mass of some 30,365,000 square kilometres which is large enough to encompass the North American continent (with a land mass of 24,230,000 square kilometres in comparison) (Britannica, 2017) and comprises 56 countries (United Nations, 2014a). Sub-Saharan Africa is a geographical term commonly used to refer to that part of the African continent that is fully or partially located south of the Sahara Desert, and it is made up of 50 countries. It excludes six Northern African countries with the exception of Sudan, which is included

in Sub-Saharan Africa (United Nations, 2014a; United Nations, 2014c; Library of Congress, 2010).

In terms of health, studies on health care and burden of disease have consistently shown that Sub-Saharan Africa lags behind the rest of the world (Murray and Lopez, 1997; World Health Organisation, 2006; Kassebaum et al., 2014; Murray et al., 2014; Wang et al., 2014). The reasons for this poor state of health are complex and often relate to social, economic, political and historic factors which have an effect on the provision of health care. The governments of the respective countries on the African continent are responsible for having mechanisms in place for health care, including the infrastructure, the provision of support services, drugs and medical supplies, and importantly, having an adequate health workforce. Training institutions play an important role in the preparation of health workers for entry into the workforce, and there is a need for quality assurance in the educational process (World Health Organisation, 2006:xx-xxi). Issues regarding the health workforce and shortages of manpower, training needs of the health workforce, and in particular, medical education of doctors, are all interlinked and impact on relevant health-care issues of the region and each will be discussed further in turn.

Graduating doctors need to be certified as competent to handle the demands of health-care pertinent to the region where they work. To ensure this happens, medical curricula need to be relevant to the context in which graduates will practise both in terms of content and what students are assessed on. In the medical programme, it would seem logical for the curriculum to have relevance to the local context in order to address the problems surrounding burden of disease and health-care, with the linking of assessment to outcomes as a next logical step (Harden, 2001). There is currently little documented evidence about curricula in Sub-Saharan Africa in relation to the health-care issues and needs of the region. A recent search using some key search sites (for the period 2010 to 2017) confirmed that this is indeed the case. Perhaps this is because outlining the health-care needs of a region like Sub-Saharan Africa can be complex and there is no simple model to apply. One example of how health-care needs and priorities have been delineated is through the Millennium Development Goals.

The United Nations Millennium Declaration (United Nations, 2000), signed at the Millennium Summit of the United Nations in September 2000 by all 189 UN member

states, committed world leaders to a global partnership to combat poverty, hunger, disease, illiteracy, environmental degradation, and discrimination against women. Following the summit, eight Millennium Development Goals (MDGs) were derived from this Declaration (United Nations, 2001:56-58). Each goal had specific targets set for 2015, with measurable indicators to monitor progress from 1990 levels (United Nations, 2008), and all are inter-dependent:

MDG 1: To eradicate extreme poverty and hunger
MDG 2: To achieve universal primary education
MDG 3: To promote gender equality and empower women
MDG 4: To reduce child mortality
MDG 5: To improve maternal health
MDG 6: To combat HIV/AIDS, malaria, and other diseases
MDG 7: To ensure environmental sustainability
MDG 8: To develop a global partnership for development

Although the MDGs have since been superseded in 2016 by the Sustainable Development Goals (SDGs) with an expansion from eight to 17 goals (Sachs, 2012; United Nations, 2016), the MDGs were applicable at the time of planning and data collection of this study and selected as one possible reference source. The MDGs have been used as a common framework for international development by economically advanced and developing countries (Hulme, 2009:47). Of the eight MDGs, three are of direct relevance to health-care, namely MDGs 4, 5 and 6. Health professionals, including doctors, would need to have appropriate training and to achieve a level of competence that would enable them to contribute to meeting these MDGs. It would therefore be plausible to use these health-related MDGs to inform the appropriate competence of doctors in Sub-Saharan Africa. The validity of exit-level assessment could be investigated with respect to the MDGs.

MDGs 4 and 5 relate primarily to the disciplines of Paediatrics (or Child Health) and Obstetrics & Gynaecology (or Maternal Health) respectively in clinical medicine. MDG 6 has a more complex relationship involving more than one clinical discipline, by nature of the conditions mentioned, with two targets of this goal relating to “HIV/AIDS” and “malaria”. However, this MDG also encompasses “other major diseases”, but only lists “tuberculosis” among its measurable indicators in addition to HIV/AIDS and malaria

(The rationale underlying this situation is unclear but will not be explored further here). As will become evident later in Section 4.3.2 a decision was made to focus on the disciplines of Paediatrics and Obstetrics & Gynaecology for this study, and MDGs 4 and 5 informed this decision.

The targets and indicators for MDGs 4, 5 and 6 are summarised in Table 3.1 with the targets for each MDG in the two left-hand columns, and the corresponding indicators for the respective MDG in the two right-hand columns. These targets and indicators were set up by consensus of experts drawn primarily from the United Nations secretariat, the International Monetary Fund, the Organisation for Economic Cooperation and Development and the World Bank (UN Millennium Project, 2005).

Table 3.1 Targets and Indicators for Millennium Development Goals 4, 5 and 6

Millennium Development Goal (MDG) 4: To reduce child mortality			
Target 4.A	To reduce by two thirds, between 1990 and 2015, the under-five mortality rate	Indicators	4.1 Under-five mortality rate 4.2 Infant mortality rate 4.3 Proportion of 1-year-old children immunised against measles
Millennium Development Goal (MDG) 5: To improve maternal health			
Target 5.A	To reduce by three quarters, between 1990 and 2015, the maternal mortality ratio	Indicators	5.1 Maternal mortality ratio 5.2 Proportion of births attended by a skilled health personnel 5.3 Contraceptive prevalence rate 5.4 Adolescent birth rate 5.5 Antenatal care coverage (Attended at least once by skilled health-care personnel and at least 4 antenatal care visits) 5.6 Unmet need for family planning
Target 5.B	To achieve, by 2015, universal access to reproductive health		
Millennium Development Goal (MDG) 6: To combat HIV/AIDS, malaria and other diseases			
Target 6.A	To have halted by 2015 and begun to reverse the spread of HIV/AIDS	Indicators	6.1 HIV prevalence among population aged 15-24 years 6.2 Condom use at least in high-risk sex 6.3 Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS 6.4 Ratio of school attendance of orphans to school attendance of non-orphans aged 10-14 years 6.5 Proportion of population with advanced HIV infection with access to
Target 6.B	To achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it		
Target 6.C	To have halted by 2015 and begun to reverse the incidence of malaria and		

	other major diseases		antiretroviral drugs 6.6 Incidence and death rates associated with malaria 6.7 Proportion of children under 5 sleeping under insecticide-treated bed nets 6.8 Proportion of children under 5 with fever who are treated with appropriate anti-malarial drugs 6.9 Incidence, prevalence and death rates associated with tuberculosis 6.10 Proportion of tuberculosis cases detected and cured under directly observed treatment short course

Source: Official List of MDG Indicators (United Nations, 2008)

While some countries have made impressive gains in achieving the targets in the health-related MDGs, others are falling behind. Often the countries making the least progress are those affected by high levels of HIV/AIDS, economic hardship or conflict, with the health-care workforce shortages contributing to this state of affairs, as is the case in Sub-Saharan Africa (World Health Organization, 2012). The more recent reports from the World Health Organisation and the United Nations on the MDGs provide some indication of the magnitude of the problems that still exist in Sub-Saharan Africa (World Health Organization, 2012; United Nations, 2013; United Nations, 2014b). Since 1990, the childhood mortality rate worldwide has dropped by 41%. However, 46% of global child deaths occur in Africa, and most are preventable. In Sub-Saharan Africa, the under-five mortality rate has fallen from 178 deaths per 1000 live births in 1990 to 109 deaths per 1000 live births in 2011. 1 in 9 children die before the age of five in Sub-Saharan Africa, more than 16 times the average for developed regions in the rest of the world. Of 50 countries in Sub-Saharan Africa, only eight are expected to achieve the MDG 4 target if current trends continue. With regard to MDG 5, 52% of global maternal deaths occur in the African Region. According to estimates of trends in maternal mortality between 1990 and 2010, the WHO African Region still has the highest maternal mortality ratio (MMR) at 480 deaths per 100,000 live births. Although substantial progress has been made in many areas, much effort is still needed in order to achieve some targets.

The MDGs each with specific targets and measureable indicators, have provided a useful framework to inform overall global health policies which, in turn, ultimately guide

health-care at a clinical level. However, there are many steps in this process, involving cooperation and collaboration among various health-care disciplines. One of the steps, as already mentioned, would require the alignment of the education of health-care professionals with the local health-care needs of the region (Frenk et al., 2010; Celletti et al., 2011; Miller et al., 2011). As will become evident in Chapter 4 Section 4.6.2 the MDGs are specified at too general a level to inform medical curricula and so need to be "unpacked" to inform what is actually taught in the curriculum and assessed. This raised the question as to how this could be achieved. Consideration therefore had to be given to alternative measures of local health-care issues.

One such alternative that was considered was drawn from the Global Burden of Disease (GBD) Study, a collaboration between the World Bank and World Health Organisation. Initiated in 1992, this study sought to quantify the global burden of premature death, disease, and injury with a measure that could be used for cost-effectiveness analysis in order make recommendations that would improve health, particularly in developing nations (Murray and Lopez, 1997; Gold et al., 2002). Analysis of the results of the GBD were by means of various epidemiological and demographic indicators, one of them being disability-adjusted life years (DALYs), which allows for the comparison of the burden of premature mortality with non-fatal health outcomes, such as disability (Murray and Lopez, 1997:1270). DALYs for a disease or health condition are calculated as the sum of the Years of Life Lost (YLL) due to premature mortality in the population and the Years Lost due to Disability (YLD) for people living with the health condition or its consequences (World Health Organisation, 2015).

In the Global Burden of Diseases, Injuries and Risk Factors Study 2010 (GBD 2010), DALYs were calculated using collected data for 291 diseases and injuries in 187 countries (Murray et al., 2012). Table 3.2 illustrates the 20 leading causes of DALYs globally and for each of the four WHO-defined Sub-Saharan African regions, with the causes of DALYs associated with MDGs 4, 5 and 6 highlighted. As indicated in the table, the corresponding rankings of individual leading causes of DALYs varied noticeably in each Sub-Saharan African region in relation to the overall global ranking, emphasizing differences which may be unique to individual regions (For example, protein energy malnutrition, a leading cause of under-five mortality rate in children, ranks within the 20 leading causes in the eastern, central and western Sub-Saharan African regions, but

not in the southern region). Eleven of the 20 leading global causes listed are due to MDG-associated conditions. Although meningitis and maternal disorders are not listed in the 20 leading global causes, they feature in the 20 leading causes for three out of four of the Sub-Saharan African regions.

Table 3.2 20 Leading Causes of Disability Adjusted Life Years (DALYs) in each of four Sub-Saharan African (SSA) Regions

Global Ranking	Cause of Disability Adjusted Life Years (DALYs)	Ranking in Sub-Saharan African regions			
		Southern SSA	Eastern SSA	Central SSA	Western SSA
1	Ischaemic heart disease	14	21	19	20
2	(#) Lower respiratory infections	2	3	4	2
3	Cerebrovascular disease	7	16	14	16
4	(#) Diarrhoeal disease	3	4	2	3
5	(# *) HIV/AIDS	1	1	5	4
6	Low back pain	15	17	23	13
7	(#) Malaria	20	2	1	1
8	(#) Preterm birth complications	6	5	6	7
9	Chronic obstructive pulmonary disease	9	20	20	22
10	Road injury	13	11	12	9
11	Major depressive disorder	10	13	17	19
12	(#) Neonatal encephalopathy (includes birth asphyxia/trauma)	12	9	10	10
13	(# *) Tuberculosis	4	7	7	12
14	Diabetes mellitus	8	29	28	26
15	(*) Iron deficiency anaemia	11	12	11	11
16	(#) Sepsis and other infectious disorders of the newborn baby	29	8	13	5
17	(#) Congenital anomalies	17	18	8	18
18	Self-harm	27	32	37	69
19	Falls	43	33	32	21
20	(#) Protein energy malnutrition	36	6	3	6
25	(#) Meningitis	24	10	9	8
27	Interpersonal violence	5	23	21	30
31	Drug use disorders	18	39	49	47
34	Fire, heat, hot substances	33	22	24	17
36	Epilepsy	19	19	25	14
40	(*) Maternal disorders	28	14	15	15
48	Exposure to mechanical forces	16	38	30	53
55	Syphilis	41	15	16	23
71	Sickle cell disorders	134	67	18	24

Key: (#) Cause of DALY associated with MDG 4 (*) Cause of DALY associated with MDG 5	Note: Causes of DALYs outside the globally ranked top 20 were included if they featured among the top 20 in any of the four listed Sub-Saharan African regions

Source: Murray et al. (2012)

It is evident that the conditions listed in Table 3.2 have a significant impact on mortality and morbidity, and it would therefore seem reasonable to expect the leading 20 causes of DALYs to be covered in the training programmes of health-care professionals, especially doctors. Causes of DALYs could provide a more tangible framework linking with clinical conditions that doctors would be expected to encounter upon graduation, whereas MDGs are situated at a more overarching level guiding policies on health-care. Hence it could be argued that these leading 20 causes of DALYs could be used as a framework with which to further explore the curriculum and assessment. Such application is to have relevance for this study.

3.2 Health-Care in Sub-Saharan Africa – Human Resources

Apart from the high burden of disease described previously, Africa has very low levels of resources – both human and financial – for health care compared to other world regions. Kinfu et al (2009) estimated the density of physicians per 1,000 population in 12 Sub-Saharan African countries studied ranged from 0.02 to 0.29. In comparison, for example, the figures for China are 18.9 physicians per 1,000 population, Brazil 14.6, United Kingdom 2.79, and 2.45 in the United States of America (World Health Organisation, 2014) (See also Figure 3.1). Data in terms of infrastructure are largely incomplete for a number of African countries, but where available, indicate equally low figures in comparison with other regions in the world. For example, the number of hospital beds per 10,000 population ranges from 5 in Uganda to 7 in Mozambique, 8 in Liberia and 18 in Botswana. In contrast, figures for Brazil are 23 per 10,000 population, for China, Norway, United Kingdom and the United State of America range from 29 to 38.

The 2006 World Health Organisation (WHO) World Health Report focussing on the global health workforce (World Health Organisation, 2006) noted that the African Region had just 1% of the total global funding for health, and only 2.3 health workers per 1,000 population (compared with 18.9 in Europe and 24.8 in the Americas) dealing with 24% of the global disease burden (World Health Organisation, 2006:5,8,9). Put in

another way, the number of doctors and nurses for the population in Sub-Saharan Africa is one tenth the number in Europe (Chen et al., 2004). In contrast, the WHO Region of the Americas had 10% of the global disease burden, and 37% of the world's health workers spending more than 50% of the world's health financing (World Health Organisation, 2006:8,9). It is therefore evident that there is a shortfall of health workers in Africa.

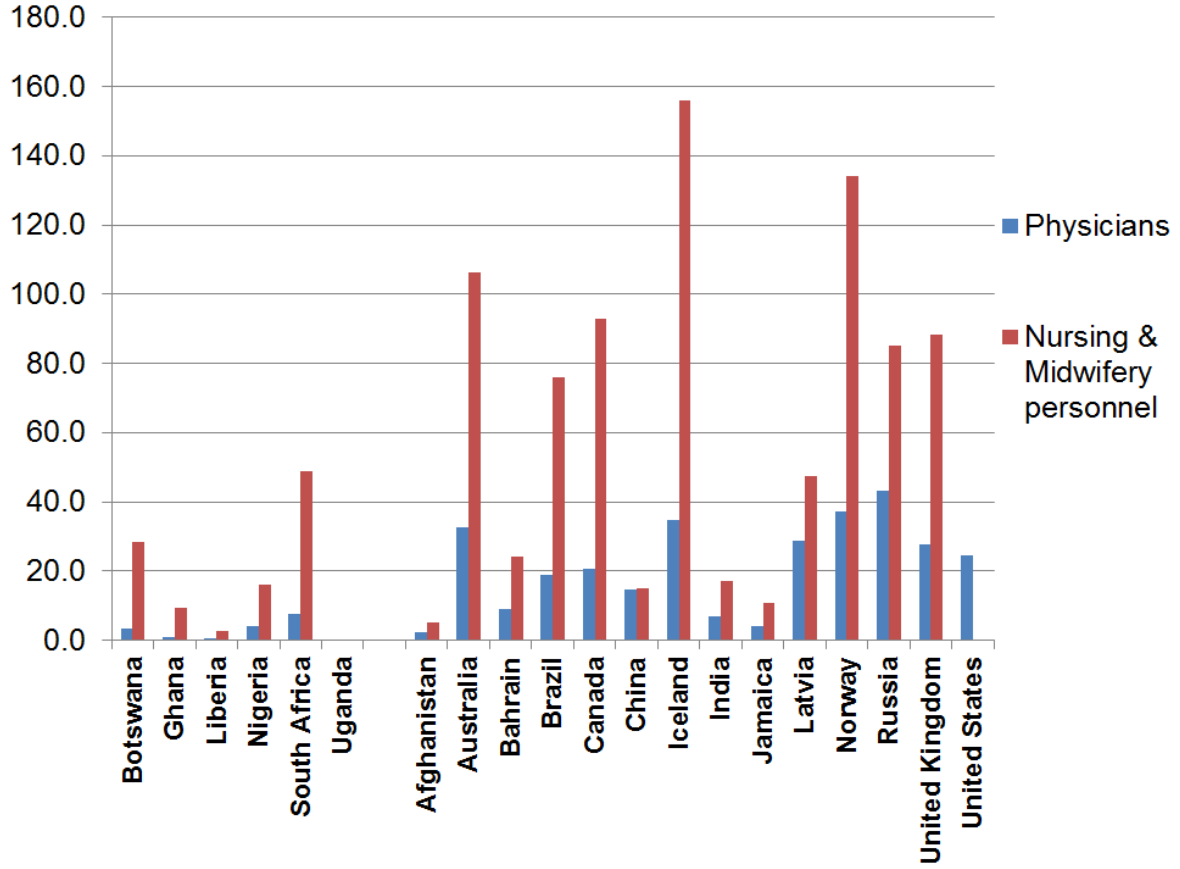


Figure 3.1 Density of the health workforce in selected United Nations member countries

Key:
Density of the health workforce (per 10,000 population) is shown for the period 2006-2013.
Countries in Sub-Saharan Africa are shown on the left side of the chart. Countries from other continents are shown on the right side.

Note: Data was not available for some countries, such as Uganda, highlighting challenges and issues of data collection.

Source: World Health Organisation (2014)

The calculated number of health-care professionals (counting only doctors, nurses and midwives) per 1,000 population that would be required in order to achieve an 80% coverage rate for deliveries by skilled birth attendants or for measles immunisation is 2.5 (Chen et al., 2004). Currently, 54 countries in the world are below this minimum threshold, and 43 of these countries are in Sub-Saharan Africa (World Health Organisation, 2014). Figure 3.1 illustrates the health workforce provision for doctors, nursing and midwifery personnel for the period 2006-2013 in selected Sub-Saharan African countries together with a selection of some other countries in the rest of the world to provide some comparison.

The factors associated with such a disparity in the numbers in different regions of the world are complex, and often include a combination of social, economic and political factors, most notably with civil wars and other conflicts playing a significant part (Hagopian et al., 2004; Gukas, 2007). A better life and livelihood are plausible reasons for the decisions to migrate, whether from rural to urban areas within countries, from poorer to better-off countries within regions and across continents (Astor et al., 2005; World Health Organisation, 2006; Simoens and Hurst, 2006; Buchan et al., 2008). In the Sub-Saharan African Medical Schools survey (SAMSS), the very low physician-to-population ratios in Sub-Saharan African countries were attributed to emigration of graduates to other countries or continents (Mullan et al., 2011:1113), and to the small numbers of doctors graduating from some of the medical schools.

3.3 Medical Schools in Sub-Saharan Africa

In the 19th and early 20th centuries, the majority of the Sub-Saharan African countries were colonised by European powers. Universities and medical schools that were established in some of the countries were modelled on the colonial institutions of the occupying colonial powers (Omer, 2005; Gukas, 2007; Monekosso, 2014).

The number of medical schools has increased rapidly especially in the second half of the 20th century. Determining precise numbers remains a challenge, however, as comprehensive information is not readily available or necessarily up to date (Karle, 2008). A comparison of data obtained from three sources illustrates this point (see Table 3.3). These sources are the Sub-Saharan African Medical Schools Survey (SAMSS) (Chen et al., 2012) and two web-based global directories of medical schools: the

International Medical Education Directory (IMED) (FAIMER, 2011) and the Institute for International Medical Education (IIME) database (Institute for International Medical Education, 2006).

The Sub-Saharan African Medical Schools survey (SAMSS) was conducted over 2008-2009 to provide a database regarding the status of medical schools and trends within medical education in this region (Mullan et al., 2011; Chen et al., 2012). The list of medical schools was compiled using a variety of sources including publicly available global directories of medical schools, databases of several international health organisations, universities and national medical professional registration bodies as well as Ministries of Health (Chen et al., 2012:2-3).

Table 3.3 Medical Schools in Sub-Saharan Africa

	IIME (2006)	SAMSS (2008-09)	IMED (2011)		IIME (2006)	SAMSS (2008-09)	IMED (2011)
Angola	1	7	1	Liberia	1	1	1
Benin	1	2	1	Madagascar	2	2	3
Botswana	-	1	1	Malawi	1	1	1
Burkina Faso	1	2	1	Mali	1	2	1
Burundi	1	1	1	Mauritania	1	1	-
Cameroon	1	4	3	Mauritius	1	2	1
Central African Republic	1	1	1	Mozambique	1	4	2
Chad	1	1	1	Namibia	-	1	-
Comoros	-	1	1	Niger	1	1	1
Congo	1	1	1	Nigeria	16	25	26
Côte d'Ivoire	2	2	2	Rwanda	1	1	1
Congo, Democratic Republic	4	19	9	Senegal	1	4	1
Djibouti	-	1	-	Seychelles	-	1	1
Equatorial Guinea	-	1	-	Sierra Leone	1	1	1
Eritrea	-	1	-	Somalia	2	2	2
Ethiopia	3	12	8	South Africa	8	8	10
Gabon	1	1	1	Sudan	14	29	27
Gambia, The	-	1	2	Tanzania	4	5	5
Ghana	2	4	2	Togo	1	1	1
Guinea	1	3	1	Uganda	3	4	5
Guinea-Bissau	1	1	1	Zambia	1	1	1
Kenya	2	3	2	Zimbabwe	1	2	1
				TOTAL	**86**	**169**	**132**

Sources: Institute for International Medical Education (IIME) (2006)
Sub-Saharan African Medical Schools Survey (SAMSS) (Mullan et al., 2011; Chen et al., 2012)
International Medical Education Directory (IMED) (2011)

The International Medical Education Directory (IMED) was developed by the Foundation for the Advancement of International Medical Education and Research (FAIMER), a non-profit foundation established in 2000 by the Educational Commission for Foreign Medical Graduates in the United States of America (Boulet et al., 2007; FAIMER, 2011). Medical schools were only listed in this directory after FAIMER received confirmation from the appropriate government agency (in most countries, this is usually the Ministry of Health) that the school is recognised (Boulet et al., 2007:21). Since 2014, IMED has been replaced by the World Directory of Medical Schools (WDMS), a collaboration between FAIMER and the World Federation of Medical Education (World Directory of Medical Schools, 2014).

The Institute for International Medical Education (IIME) was established in 1999 by the China Medical Board to address global issues of medical education, including quality standards. The information for the IIME reference database came from a medical school survey conducted in 2000 (Institute for International Medical Education, 2006).

The data provided by these sources is not without its problems as is evident from the discrepancies in the numbers of medical schools listed for various countries that are unlikely to merely represent changes over time. As an example, the IMED database listed ten medical schools in South Africa. From the South African perspective, this is clearly incorrect as there were only eight medical schools in the country in 2011. On accessing the website, it emerged that one university has been listed twice (as the University of Transkei and Walter Sisulu University), although this would not be immediately obvious unless the reader or researcher was aware of the historical background of this institution and its transition from the former into the latter name. The WDMS database has since indicated that these two university names refer to the same institution (although the number of medical schools in South Africa is still listed as ten), but nevertheless, this example serves to highlight the challenges of maintaining accuracy of information with databases particularly in under-resourced contexts. Another issue to bear in mind that the term “medical school” is used loosely and cannot be assumed to relate solely to the training of doctors, as in some countries it may encompass facilities for teaching other health-care professionals such as nurses, pharmacists and other allied health workers (Chen et al., 2012).

Notwithstanding issues with the quality of available data, the reality is that while some countries have multiple medical schools, there are a number of Sub-Saharan African countries that do not have any medical schools. In some countries, the number of medical doctors graduating annually from medical schools is low in proportion to the respective country population, contributing to an acute shortage of doctors and other trained health workers in Sub-Saharan Africa (Chen et al., 2004; World Health Organisation, 2006; Anyangwe and Mtonga, 2007; Mullan and Frehywot, 2008; Kinfu et al., 2009). In looking at the challenges and emerging trends in medical education in Sub-Saharan Africa, workforce shortages were also noted in the SAMSS study (Mullan et al., 2011). As a stable and sufficient health-care workforce is essential to meet the health-care needs of any population (Chen et al., 2012), the deficiencies in these resources will inevitably have an impact on the health-care needs of the population. The lack of trained doctors coupled with high clinical workloads and working in academic institutions where teaching does not have the same status or priority as research could also impact significantly on the teaching that takes place, what is assessed and how the assessment is conducted (Frenk et al., 2010).

As will be seen in later sections in this study, challenges relating to resources and a sufficient health-care workforce have particular relevance on the training provided in medical schools.

3.4 Assessment of clinical competence in a resource-constrained context such as Sub-Saharan Africa

Given that medical schools must be staffed by trained and experienced professionals, it stands to reason that if countries have very low numbers of healthcare workers, those countries will also struggle to resource their training facilities adequately. Modest salaries, limited career options, heavy teaching loads, growing enrolment and absence of equipment and support staff have been the main barriers to retention of faculty (Mullan and Frehywot, 2008; Mullan et al., 2011; Mullan et al., 2012; Talib et al., 2013; Olapade-Olaopa et al., 2014). Migration has already been mentioned earlier in this chapter as a possible factor for the 'brain drain' leading to a critical shortage of trained health-care workers in Sub-Saharan Africa (Hagopian et al., 2004; Astor et al., 2005; World Health Organisation, 2006). Although socioeconomic reasons may be at the forefront of decisions to migrate, consideration should also be given to the possibility

that if the training received by doctors in a medical school is modelled upon curricula from Western countries that formerly colonised the African continent, this could make it more conducive for graduates to emigrate towards those Western countries as their training may facilitate their employment prospects over there.

In addition to teaching, medical school faculty are also responsible for assessment activities, including exit-level assessment that leads to registration. There is currently no available information regarding national registration examinations in Africa which suggests that the responsibility for certification of competence typically lies with the relevant institution. Assessment is typically a resource-intensive activity. This is more so for assessments of clinical competence, as these are more intensive assessments that have to be planned and conducted by clinicians who are already juggling heavy clinical workloads. The ability to mobilise needed resources to design and mount assessments of clinical competence will impact directly on the quality of those assessments.

The focus of the Sub-Saharan African Medical Schools (SAMSS) Study provided baseline data on the medical schools including infrastructure and capacity (Mullan et al., 2011; Chen et al., 2012). However, these reports do not address current assessment practices in medical schools on the continent, nor do they describe what measures are taken in these institutions to ensure that these assessments are appropriate in this context. Information on curriculum issues and assessment practices in particular (apart from curricular innovations such as problem-based learning and community-based learning) is largely missing in Greysen et al's (2011) review of existing literature on medical education coming from Sub-Saharan Africa, although there has been some work published regarding clinical assessment of medical students on the continent. The following paragraph provides some examples of such published work.

Burch and Seggie (2008) reported on the use of portfolio assessment using a structured interview in a resource-constrained setting, which has since been used in other institutions (Cameron et al., 2011; van Schalkwyk et al., 2012). There are several short reports on the use of Objective Structured Clinical Examinations (Oviasu, 2008; Odoi Adome and Kitutu, 2008), and the use of oral examinations has been described in surgery and psychiatry (Stupart et al., 2008; Niehaus et al., 2012). More recent research has included the use of tools such as an online platform to help prepare students for summative assessment (Kerr et al., 2016) and to assess reflection (Wium and du Plessis,

2016). The application of a simple method of standard setting in assessment has also been considered (Wright, 2016). There are no published studies providing a summary of exit-level assessment practices in Sub-Saharan African medical schools comparable to McCrorie and Boursicot's survey (2009) of British medical schools, and the survey conducted on Australian medical schools by Ingham (2011).

A search of one other pertinent online source for literature on any issues relating to assessment in Sub-Saharan Africa yielded no results (African Journals Online, 2015). International conferences serve as an alternative platform for the dissemination of research conducted in medical education. A review of the conference abstract books over the period 2001 to 2014 for the annual Association for Medical Education in Europe conferences (AMEE, 2015) revealed 8 short communications and 9 poster presentations on the subject of clinical assessment in Sub-Saharan Africa, all originating from South Africa. A similar review of available conference abstract for the Ottawa Conference on the Assessment of Competence in Medicine and the Healthcare Professions (OTTAWA, 2015) for past conferences in 2010, 2012 and 2014 yielded a total of 2 short communications and 6 poster presentations on the subject of clinical assessment in Sub-Saharan Africa. In total, only 5 short communications and 4 poster presentations from these two major international conferences addressed exit-level assessment issues that were relevant to this thesis (Three related to assessment methods, two to assessment mapping, and one each on changing assessment practice, assessment performance, practical skills and students' perceptions on competence).

One multinational African collaboration, already discussed in Section 2.7, did focus on assessment-devised guidelines on the selection of sustainable assessment methods in the limited resource settings of most African medical schools (Walubo et al., 2003). Their model scored and weighted selected assessment methods against performance, cost, suitability and safety factors, and underlined the dilemma of choice of assessment methods based on sound assessment practices and the limit on resources needed to conduct these practices.

This brings us to an issue of much importance for assessment in medical education in Sub-Saharan Africa, namely, what then, should medical schools in Africa be striving for to certify their graduates as competent? In order to answer this question perhaps it is necessary to consider several further questions with regard to assessment: (1) What

exit-level assessments take place in Sub-Saharan African medical schools, where there are significant resource constraints?, and more importantly, (2) Are the exit-level assessments that are currently in place appropriate both in terms of what they are measuring and how they are measuring the competence of graduating students? As a consequence, striving to answer these questions on assessment in a Sub-Saharan African context led to a single-site study that was to serve as a pilot for the work reported on in this thesis.

3.5 Preliminary study on assessment of clinical competence in one Sub-Saharan African institution

In light of the paucity of published information about assessment practices in Sub-Saharan Africa, a starting point was to determine what assessment was presently taking place in one Sub-Saharan African medical school. A preliminary study was carried out in April and May 2013, to map current assessment practice in the final 18-month phase of the Bachelor of Medicine & Bachelor of Surgery (MB,ChB) programme at Stellenbosch University, South Africa, based on the way in which assessment methods are described in the official faculty documents and clinical clerkship study guides provided to the students (Tan et al., 2016)(See also Appendix A). Prior to this study, no overall map existed of assessments as practised during this period, and creating such a map would help provide an overall picture of what assessment takes place.

This descriptive interpretive study centred on a process of document analysis of the 2012/2013 study guides for the final 18-month phase of the six-year MB,ChB (undergraduate medical) programme at Stellenbosch University. During this phase, students rotate through 11 clinical clerkships varying in length from three to seven weeks. Detailed information relating to each clinical clerkship, including teaching schedules, duty rosters, projects and assignments, assessment methods and resource materials, is made available in the study guides that are provided to all students and relevant faculty. The information was collated on an Excel spreadsheet and categorised by clinical clerkships and assessment methods to generate an overview of exit-level assessment in the programme and to facilitate comparison between the clinical clerkships. As this process proceeded, it became clear there were some gaps in information available from the study guides. Clarificatory interviews were therefore sought with clinical clerkship coordinators to verify and add to the correctness of

assessment-related information in the study guides. An existing framework for the assessment of clinical competence (Miller, 1990), also known as Miller's 'pyramid' (as described in Chapter 2), was used in the analysis of the data obtained.

3.6.1 Findings of the Preliminary Study

Methods used appeared appropriate to exit-level, with a range of assessment methods especially in the end of clerkship assessments. In the final exit assessment, the methods used seemed to indicate a predominance of performance-based assessment (which would correspond with the "Shows How" tier of Miller's 'pyramid'), but as similar assessment methods used were described differently across different clinical clerkships it is unclear whether the similar assessment methods were actually one and the same. Questions also arose regarding the application of different weightings by different clinical clerkships to similar assessment methods and why this should be the case. While providing an in-depth analysis of assessment methods across an undergraduate medical programme, this preliminary study has called into question the strength of decisions about graduating doctors.

In addition, the preliminary study highlighted a number of methodological considerations for later work. These included the four key findings that emerged from this preliminary study (Tan et al., 2016). Firstly, there was a diversity of methods and approaches to assessment across modules in the final phase clinical clerkships. Secondly, clinical clerkships using similar assessment methods weighted the contribution of those methods differently during scoring. A third finding was that the information provided about similar assessment methods was described differently in the various clinical clerkship study guides for students. These do not necessarily have the same meaning nor are they synonymous with what is described in the literature. And fourthly, study guides varied in the amount and detail of information provided about the assessment methods used in the respective clinical clerkships.

This preliminary study therefore highlighted potential areas where current practice needed to be investigated in greater depth, and provided the impetus for further investigation into exit-level assessment in a resource-constrained setting. As an example, there appeared to be factors at play that could influence exit-level assessment practices, such as whether resource constraints are taken into consideration when selecting assessment methods, as well as how they influence (whether consciously or

unconsciously) the decisions that are made. Exploring reasons for choices of assessment methods was not part of the preliminary study, and could also be considered for further study.

Based on what has been described thus far in this thesis, a conceptual framework was developed in an attempt to illustrate the complex interplay between education and health of a community, and which has helped lead to the setting for this research study. This will be elaborated further in the following Section 3.6.

3.6 A Conceptual Framework

During the course of exploring and trying to gain some understanding of the health issues and burden of disease that exists in the Sub-Saharan African region it became apparent that the issue of "alignment" was brought up several times. As noted earlier in Section 2.1.1 with constructive alignment (Biggs, 1996), attention was drawn to the importance that assessment – particularly at exit-level – should be aligned with the intended outcomes of curricula. Those curricula should, in turn, be aligned with the health-care needs of the country. In discussing the preparation of the health workforce, the 2006 World Health Report noted that the curriculum is "expected to meet the standards that are often defined as core competencies", and that careful attention needs to be given to "aligning what is being taught to what is appropriate…" (World Health Organisation, 2006:47-48). And while drawing attention to the divide between the education of health professionals and the health-care systems of countries, Frenk et al. (2010) have provided an outline showing the connections between the education and health systems, and call for "new instructional and institutional strategies" to be designed so that health-care professionals can deal with the rising challenges of health-care in the 21st century. The mismatch of competencies to patient and population requirements highlighted by Frenk et al. (2010) needs to be addressed, a sentiment echoed by Celletti et al. (2011) and Miller et al. (2011) when proposing the alignment of health sciences education programmes with the health-care needs of the country, especially in countries that are developing or have resource-constraints, in order to facilitate efforts to improve health-care. A conceptual framework was therefore generated to try and draw together all these key factors and alignments.

Miles and Huberman (1994:20) defined a conceptual framework as "the current version of the researcher's map of the territory being investigated". Leshem and Trafford explored conceptual frameworks further, drawing upon the thinking of different authors, and surmised that conceptual frameworks could provide "a bridge between paradigms which explain the research issue and the practice of investigating that issue" (Leshem and Trafford, 2007:99). Bordage (2009:315) used lighthouses and lenses as analogies to explain how conceptual frameworks can help understand (lighthouse to illuminate) problems and emphasize (lens to magnify) aspects of the problems or elements of the solutions. Bordage's description appears to best fit the conceptual framework for this study that is presented here.

With reference to Figure 3.2, the conceptual framework illustrates the relationship between education and health of the community in the Sub-Saharan African context.

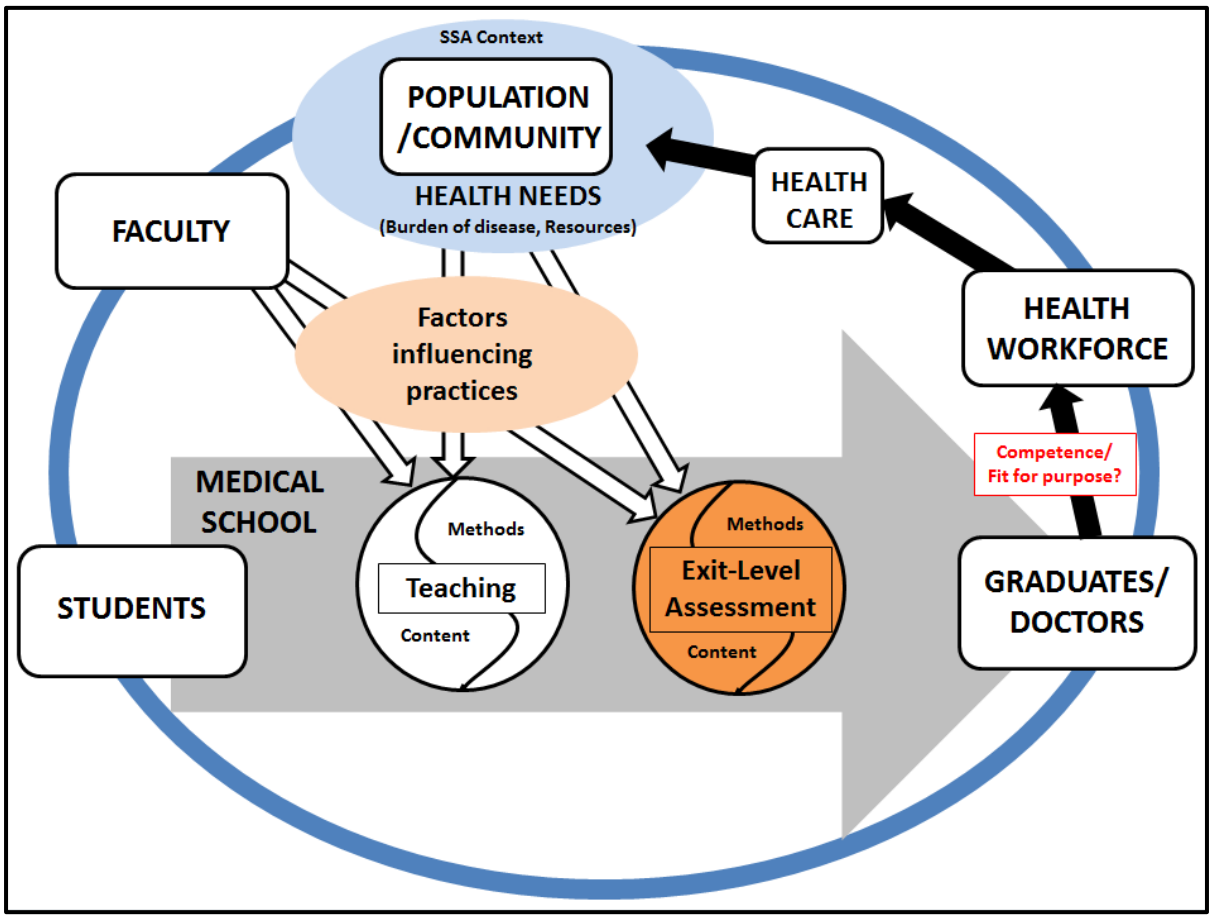


Figure 3.2 A Conceptual Framework

The outer blue ring links the main players in the health-care of this community, namely, the population, faculty, students, graduates and health workforce. The shaded grey arrow shows the direction of students moving through a medical school programme encompassing assessment as well as teaching and learning taking place throughout the course of their training. The solid black arrows denote "contribution to", while the white arrows signify "exert an influence on". Upon graduation the newly-qualified doctors enter the health workforce to provide health-care to the community. The health needs of the community are shaped by the existing burden of disease and available resources, which in turn (should) have an influence on teaching and assessment in the curriculum. Similarly, the medical school faculty who teach the students are also involved in assessing them, and their practices in both these areas may well be influenced by the health-care needs of the community.

In aligning education with health, it appears these will both be affected if the clinical competence of doctors is in question and they are not fit for purpose upon graduation from medical school. This aspect requires further in-depth investigation and leads to the aim of this study focussing on exit-level assessment. This will be elaborated further in the next chapter, together with the research methodology for this study.

Chapter 4

Methodology

Overview of Chapter

The previous two chapters focussed on theoretical perspectives and the context of this study. This chapter focuses on the research methodology of the study and what was done with regard to approaches and strategies to achieve its aim. Starting with the purpose and aim of the study, the research question is defined and the objectives of the study are outlined. This leads on to the research approach and research design that were chosen, followed by descriptions of the sample settings. Measures towards research integrity (including ethical considerations) and steps taken to ensure quality of the data, followed by the various data collection methods that were utilised and the approach taken to analyse the data collected bring this chapter to a close.

4.1 Introduction

The preliminary study on assessment methods described in the previous chapter highlighted potential areas where current practice needed to be investigated in greater depth. That study, conducted in one Sub-Saharan African undergraduate medical programme, led to the contemplation of whether and how findings would differ in similar resource-constrained settings in the African continent.

4.1.1 Purpose and Aim of the Study

As noted above, the poor state of health-care in Sub-Saharan Africa in comparison to many other regions in the world has been attributed primarily to the huge burden of disease together with a shortage of trained health-care professionals (World Health Organisation, 2006). Aligning the training – both teaching and assessment – of health-care professionals, including doctors, with the health-care needs and priorities of the community could lead to graduates who are better equipped to manage the challenges of the context in which they will be working.

The aim of this study was, therefore, to explore the validity of exit-level assessments in relation to the clinical competence required of medical graduates in resource-constrained settings in Sub-Saharan Africa to deliver selected health-related issues (The health-related issues will become apparent later in Section 4.3.2).

4.1.2 Research Question

The research question, forming the crux of this study, was:

How valid are the exit-level assessment (ELA) practices in selected medical schools in Africa in determining the clinical competence of medical graduates for selected health-related issues in the Sub-Saharan African context?

4.1.3 Objectives

The objectives of the study, conducted in selected Sub-Saharan African medical schools, were:

- To investigate what ELA methods were stated as being used in resource-constrained settings
- To investigate what ELA methods were actually being used in resource-constrained settings
- To explore the reasons as to why these methods were chosen
- To investigate what clinical competencies were considered appropriate for medical graduates from these medical schools in order to address selected health-related issues

4.2 Methodology of Research

In selecting a research methodology, a researcher is confronted by an array of options that range from methodology appropriate to the "hard sciences", to those appropriate to the "soft sciences" (Biglan, 1973; Labaree, 1998). The so-called "hard sciences" comprise physical sciences, natural sciences and engineering, and are often associated with quantitative-research methods, whereas the "soft sciences", which include social sciences, education and humanities, are more associated with qualitative-research methods. The approaches in quantitative research methodology are used to test a hypothesis and to answer questions about 'how much' or 'how many', with quantitative methods generating numerical data. Qualitative-research methodology approaches, on

the other hand, are more exploratory, seeking answers to the 'what', 'how' or 'why' of a phenomenon, looking at attitudes or behaviours of individuals being studied, and language data (written or oral) is generated by qualitative methods (Malterud, 2001a; Cleland et al., 2013). Although the "hard" and "soft" descriptions of sciences are perhaps used less often now than they used to be in the past, they are useful in understanding why research activities can be quite different across the range of disciplinary domains such as those mentioned above.

The approach used in this research study has been guided by the research problem that has been identified and the research question posed (See Section 4.1) which forms the specific goal of the study. Denscombe (2010:3) advocates a strategy or plan of action designed to achieve this specific goal. This includes "an overview of the whole study… as the basis for deciding how to approach the research" (a research paradigm), and "a carefully constructed plan of action that is rationally designed and likely to offer the best prospects of success" (a research design).

4.2.1 Paradigmatic Concerns and Research Approach

"Paradigms" are sets of beliefs and practices that provide a general perspective to guide an understanding of the phenomenon being investigated (Weaver and Olson, 2006; Bunniss and Kelly, 2010; Tavakol and Sandars, 2014a). Paradigms have also been defined as "sets of philosophical underpinnings from which specific research approaches (e.g. qualitative or quantitative methods) flow" (Weaver and Olson, 2006:460).

This study was conducted within an interpretive paradigm, wherein "meaning is constructed in the researcher-participant interaction in the natural environment" (Babbie and Mouton, 2001; Weaver and Olson, 2006; Bunniss and Kelly, 2010). Interpretivism, as summarised by Bunniss and Kelly (2010:361), applies to this study:

- Reality is subjective and changing. There is no one ultimate truth.
- The nature of the knowledge is subjective. There are multiple, diverse interpretations of reality, and there is no one ultimate or 'correct' way of knowing.

- The focus of the research approach is on understanding. Meaning is constructed in the researcher-participant interaction. The interaction takes place in the natural environment, and uses inductive reasoning.
- Qualitative methods tend to be used to capture various interpretations of a phenomenon.

Qualitative data-gathering methods utilised for this study were considered the best fit, in order to address the research question raised and the study objectives through developing an understanding of the meaning and experiences gained, as opposed to working towards a particular hypothesis.

4.2.2 Research Design

Selecting a research design for a qualitative study is all about finding the appropriate fit based on the nature of the investigation and the purpose of the research (Denscombe, 2010). The research design selected for this study was multiple case studies. The case study lends itself to placement within an interpretive paradigm (Henning et al., 2004:40) and thus fits well with what was intended in this study. In addition, the goal or aim of this study was exploratory and descriptive in nature, which also fits the case study requirements (Denscombe, 2010:55). This research design was suitable for answering the 'what', 'why' and 'how' questions (Kyburz-Graber, 2004) posed earlier in the form of the study research question and objectives. Case studies have a number of characteristics, as described by Denscombe (2010) and these will each be discussed in turn in the remainder of this section, together with reference to this research study.

Case studies generally "focus on one (or just a few) instance(s) of a particular phenomenon or entity with a view to providing an in-depth account of events or processes occurring in that particular instance" or context, using a variety of methods (Denscombe, 2010:52). The "entity", or unit of analysis, being investigated (Babbie and Mouton, 2001:84) in this study was exit-level assessment practices in undergraduate medical curricula. A variety of methods was used to investigate the exit-level assessment practices taking place in several selected Sub-Saharan African medical schools, providing for the "thick description" (Babbie and Mouton, 2001:277) required for qualitative studies of this nature. The case study design was relevant for this particular study because the research was taking place in the natural setting and in real

life, involving complex social situations (Crowe et al., 2011; Yin, 2014). In terms of the 'natural setting', the 'case' that forms the basis of the research investigation is normally something that already exists (which in this case, would be the selected medical schools), and it not a situation that is artificially generated specifically for the purposes of research.

Case studies focus more on relationships and processes, and less on outcomes, seeking to "unravel the complexities of a given situation" (Denscombe, 2010:53) and in so doing to provide the researcher with an opportunity to view the case in its entirety rather than to hone in on a specific variable. In order to understand one thing, it is necessary to understand many others, and crucially, how the various parts are linked. It may offer an opportunity to explain why certain outcomes might happen, and this would be ideal for understanding what is actually taking place in terms of exit-level assessment practices. Denscombe's description above was particularly useful when critically reviewing the complicated matrix of data collected in this study later, serving as a reminder that this was typical of the case study process.

Multiple case studies would enable an exploration of similarities and differences between cases, offering the advantage of allowing comparisons to be made (Baxter and Jack, 2008; Yin, 2014). This was the instance here as the study was conducted at three selected medical schools in Sub-Saharan Africa (to be explained further in the following Section 4.3). However, multiple sites could easily lead to the collection of large volumes of data and if time constraints were in place, this would affect the depth of analysis and be considered a limitation.

There are some advantages and disadvantages (or limitations) to consider with the case study approach. In terms of advantages, the focus on one or a few instances allows the researcher to "deal with the subtleties and intricacies of complex social situations" (Denscombe, 2010:62). The case study approach also allows for the use of multiple methods, and encourages the use of multiple sources of data, thereby facilitating the validation of data through triangulation. This approach is particularly suitable where researcher has little control over events, and does not exert pressure on the researcher to impose controls or change circumstances. This was indeed the situation with this study, where I, as the researcher, was an outsider and had no involvement in making any decisions relating to the curriculum at any of the study sites.

While the case study approach provides an ideal vehicle for in-depth social research, there are limitations or disadvantages to this approach. Most significant of these relates to the question of generalizability of the findings that emerge from case study research. In order to address this limitation, the case study researcher needs to take care when making comparisons with other similar cases and to acknowledge that generalizability of case study findings is rather shown by demonstrating the links between findings and previous knowledge, thus providing significant opportunity for developing theory (Babbie and Mouton, 2001:283). Case studies are often "perceived as producing 'soft' data", "lacking the degree of rigour expected of social science research" (Denscombe, 2010:63). However, data produced in research of this nature has become increasingly valued (Kyburz-Graber, 2004; Baxter and Jack, 2008; Houghton et al., 2013). Issues relating to rigour in this research study will be addressed later in Section 4.8. There are some other commonly listed disadvantages of the case study including issues around obtaining access to study participants (Denscombe, 2010:63), and the 'observer' effect of the researcher (Denscombe, 2010:69) – these will be addressed as they are encountered and dealt with during the data-collection process in the following sections. Access would be an important consideration when selecting study sites, and whether the study sites were agreeable to being investigated (Crowe et al., 2011).

To summarise, this section has looked into the qualitative-research approach taken for this study. Within an interpretive paradigm, the choice of research design was multiple case studies. The next section will now focus on the study sites.

4.3 Study Sites

4.3.1 Choice of study sites

As this study was an investigation of exit-level assessment in the Sub-Saharan African context, all the available medical schools in this region were, in theory, eligible as study sites. For the purposes of the case study, it was deemed that three cases would provide opportunity to include different contexts while ensuring that the study remained feasible. Pragmatic choices were made for the selected study sites. This took into account the consideration of a purposive-sampling strategy, that is, based on criteria (Tavakol and Sandars, 2014b), and with the purpose of the study in mind (Babbie and Mouton, 2001:166).

The three medical schools in Sub-Saharan Africa selected for this study were, namely, (1) the University of Botswana School of Medicine in Botswana, (2) Stellenbosch University Faculty of Medicine and Health Sciences in South Africa, and (3) Makerere University College of Health Sciences in Uganda. Stellenbosch University was the starting point for this study because this was the site of the preliminary study to map current assessment practice in the final 18-month phase of the Bachelor of Medicine & Bachelor of Surgery (MB,ChB) programme, as described in Chapter 3.

These schools were purposively selected for a number of reasons. Firstly, they span a wide range in the number of years they have been established (Chen et al., 2012), from 8 years (University of Botswana, established 2009; its first cohort of students graduated in 2014) to mid-range at 61 years (Stellenbosch University; established 1956) and 95 years (Makerere University; established 1922, making it the longest-established in Sub-Saharan Africa) respectively. The founding of these medical schools was inexorably linked with their past colonial history and British influence in the respective countries (The Commonwealth, 2015). Secondly, they were all former British dominions, and as a result of this heritage English was widely spoken in these countries. At the time of planning it was envisaged that the fieldwork and data collection at these three sites would predominantly be carried out in this language as it was common to all three settings. Thirdly, all three medical schools were relatively easily accessible in terms of travel logistics and personal contacts through a network of health-professions educators in Botswana and Uganda also helped facilitate access at these two sites. Details of the three study sites are summarised in Table 4.1.

Table 4.1 Summary of details of the study sites

University and date of establishment	Name of medical school	Founding of medical school	Site of main campus and medical school	Main teaching hospital affiliated to medical school	Duration of medical programme	Annual intake at time of study (2014-2015)	Degree awarded
UNIVERSITY OF BOTSWANA 1982	University of Botswana School of Medicine	2005	Gaborone, BOTSWANA	Princess Marina Hospital, Gaborone	5 years	45	MBBS*
STELLENBOSCH UNIVERSITY 1866	Stellenbosch University Faculty of Medicine and Health Sciences	1956	Tygerberg, Western Cape Province, SOUTH AFRICA	Tygerberg Hospital, Tygerberg	6 years	177	MBChB*
MAKERERE UNIVERSITY 1922	Makerere University College of Health Sciences	1924	Kampala, UGANDA	Mulago Hospital, Kampala	5 years	106	MBChB*

[** Refer to text for explanation of abbreviations used*]

These three institutions were also Medical Education Partnership Initiative (MEPI) grant sites. This US-funded five-year initiative (2010-2015) aimed to support medical education and research in Sub-Saharan African institutions, with a focus on strengthening manpower resources in order to address national health needs (Medical Education Partnership Initiative, 2013). This research study was in line with one of MEPI's core themes emphasizing regionally relevant research.

University of Botswana was established in 1982, although historically its roots date back to 1964 with the formation of the University of Basutoland (Lesotho), Bechuanaland (Botswana) and Swaziland, during the time when these three countries were under British rule (University of Botswana, 2015b; The Commonwealth, 2017). The Faculty of Health Sciences was founded more recently in 2005, comprising the Schools of Allied Health Professions, Medicine, Nursing and Public Health (University of Botswana, 2015a). The Bachelor of Medicine and Bachelor of Surgery (MBBS) professional degree programme is 5 years' in duration and at undergraduate level, with students entering the programme directly from secondary-school level. The first students for the degree programme were enrolled in 2009, graduating in 2014. The main university campus and medical school are situated in the capital, Gaborone, and this is the only medical school in the country. The main teaching hospital affiliated to the medical school and also a government referral hospital is the Princess Marina Hospital, approximately 15-minutes' walking distance from the medical school.

The five-year MBBS programme at University of Botswana comprises two phases, Phase I covering the basic sciences in the first two years, followed by clinical rotations in years 3 to 5 in Phase II. The final-year rotations (each of 8 weeks' duration) are in Internal (General) Medicine, Surgery (comprising General Surgery, Anaesthesiology and Emergency Medicine), Family Medicine, Obstetrics & Gynaecology, and Paediatrics & Adolescent Health. An integrated final (exit) examination, consisting of written and clinical components, takes place at the end of the final year. The clinical disciplines, grouped either as medical-based or surgical-based disciplines, contribute a pre-determined number of questions (in proportion to total duration of teaching time allocated to the respective clinical disciplines in Phase II) for both the written and clinical components and these questions are marked by examiners from the individual disciplines.

Stellenbosch University dates back to 1866 (Stellenbosch University, 2015a). The then Faculty of Medicine only began training its first intake of doctors in 1956 (Stellenbosch University, 2015b:15). The six-year degree programme for the Bachelor of Medicine and Bachelor of Surgery (MB,ChB) is at undergraduate level, with the intake of students directly from secondary school. The present Faculty of Medicine and Health Sciences has its main campus in Tygerberg in Western Cape Province of South Africa, and is situated next to Tygerberg Hospital, a tertiary national referral hospital in the province which also serves as the main teaching hospital for the medical school. The medical school was at the time of the study one of eight medical schools in South Africa. The curriculum has clinical clerkships in the first three years based on organ systems (for example, cardiovascular system, respiratory system, digestive system and so forth), with clinical exposure starting at the end of the second year and continuing into the final year. Summative assessment in the clinical years is discipline-based. The final phase clinical clerkships ('Late Clinical Rotations') stretch over 18 months. Students rotate through eleven clerkships of varying length, from Internal Medicine and Psychiatry (7 weeks each), Obstetrics & Gynaecology and Paediatrics (6 weeks each), Orthopaedic Surgery and General Surgery (5 weeks each), Health, Disease and Disability in the Community (a five-week clerkship involving Family Medicine, Community Medicine and Rehabilitation Medicine), to Anaesthesiology, Otorhinolaryngology, Ophthalmology and Urology (3 weeks each).

Makerere University was established in 1922 (Makerere University, 2015b) and is situated in Kampala, the capital of Uganda. The then Faculty of Medicine started in 1924. With a change from a faculty-based to collegiate system around 2009, the present School of Medicine is situated within the College of Health Sciences (Makerere University, 2015a) along with the Schools of Health Sciences, Biomedical Sciences and Public Health. The Bachelor of Medicine and Bachelor of Surgery (MBChB) is a five-year undergraduate-degree programme with the intake of students from high school. Summative assessment in the clinical years at this medical school is discipline-based. In the final year comprising 2 semesters, students rotate through 5-week clinical clerkships in Clinical Medicine, Clinical Surgery, Medical Specialties and Surgical Specialties, Paediatrics & Child Health, and Obstetrics & Gynaecology. Makerere University is one of five medical schools in Uganda. Mulago Hospital, the main national

regional hospital, is situated adjacent to the School of Medicine and is a teaching hospital for the Makerere University College of Health Sciences.

A point to note is that all three medical schools are situated in urban areas, and with the adjoining teaching hospitals serving as tertiary referral hospitals, the spectrum of clinical conditions seen and treated at these hospitals would most likely be different compared to other regions of the respective countries, and would include cases of a more complex nature. While firm evidence to support this suggestion is not currently available, it is reasonable to consider this in light of what is already known about this region of the world. In Section 3.2, attention was drawn to the magnitude of the burden of disease in Sub-Saharan Africa and the lack of resources including health work-force shortages and poor infrastructure. In such a situation, the access to health-care may be more available in urban areas where larger numbers of health-care professionals have a tendency to concentrate. One consequence of reduced access to health-care for the community in more rural areas is that patients may wait longer before seeking treatment at more advanced stages of their disease conditions, which would explain the complexity of their cases.

To summarise, these three selected cases offer a study of current assessment practices taking place at three Sub-Saharan African medical schools. Two of the medical schools are well-established institutions and the remaining medical school has a more recent history with the graduation of its first medical cohort the year before this study was conducted.

4.3.2 Study Population

The study population from which the sample is drawn refers to all individuals involved in undergraduate assessment in the three selected study sites. The sample population relevant to this study focussed on individuals involved in exit-level assessments (Denscombe, 2010:23; Ritchie et al., 2003:86-87). Potential study participants were derived from two main groups, namely, faculty (including clinical clerkship coordinators and teaching faculty) who were examiners and conducted the assessments, and interns (recent graduates who underwent the assessments in order to graduate as doctors). The actual sampling process undertaken will be further explained in the following section 4.4 of this chapter.

As a proxy for the entire curriculum, Obstetrics & Gynaecology and Paediatrics were specifically chosen for several reasons. Previously in Chapter 3, the issue of whether undergraduate training in Sub-Saharan African medical schools enabled graduates to address health-care needs and priorities of the communities they will be working in was discussed. The complexity of defining those health-care needs in Sub-Saharan Africa was raised (see Section 3.1), and Millennium Development Goals (MDGs), with each goal having set targets and measurable indicators of progress, were considered as one possible framework. MDGs 4 and 5 related primarily to Paediatrics (or Child Health) and Obstetrics & Gynaecology (or Maternal Health) respectively. Although this approach ultimately proved to have limitations as a measure of health-care needs for this study (see Section 5.3.2), it provided a useful starting point for the consideration of choice of samples, and the decision was made to remain with the choice of Obstetrics & Gynaecology and Paediatrics. Additionally, these two clinical disciplines deal with pregnant women and children who pertain to vulnerable groups in emergency or disaster situations (World Health Organization, 2017) and who take a relatively high share of the burden of disease occurring in these situations. It would therefore not be unreasonable to focus some attention on such groups. Having now discussed the study sites, attention will next turn to data sources in Section 4.4.

4.4 Data Sources

In order to address the objectives of the study, the data required needed to come from a variety of sources which will be discussed further in the following subsections (4.4.1, 4.4.2 and 4.4.3), followed by an explanation of the order in which the methods were used. At each study site, it was envisaged that key individuals such as the discipline head of department or clinical clerkship coordinator would be able to provide most of the information required as they would be more likely to have an overall view of the curriculum as well as the assessment pertaining to their discipline. However, drawing upon experiences gathered previously in the preliminary study on assessment of clinical competence (Tan et al., 2016), it was considered likely that there would be no single source that could comprehensively provide all the information. Instead, the approach taken in this study was to obtain the information required from various sources and then assemble it all to provide an overview of the exit-level assessment practices taking place at that study site. Three methods were used for data collection, namely document

analysis, observations, and interviews. A summary of the methods used for data collection in relation to research study objectives and data sources is provided in Table 4.5 at the end of this Section 4.4. Detailed access to data sources was only possible for final examinations and not for in-course assessment, and as will become evident in Chapter 5, some of the findings relate only to the final examinations.

4.4.1 Document analysis

Documents can be treated as a source of data (Bowen, 2009:27; Denscombe, 2010:216) and in this case, documents in the form of written sources were analysed. In this study, document analysis was used to answer the question, "What exit-level assessment methods were stated as being used?" which forms the first study objective. Information relating to exit-level assessments was extracted from documents, which included study guides and logbooks, assessment schedules, written and clinical assessment questions, model answers, marking guides, mark sheets or checklists. This information, comprising whether the assessment was formative (continuous) or summative in nature, the assessment methods used, weightings and timing of the assessments, was then collated on an Excel spreadsheet. A separate Excel spreadsheet was used for each study site. Table 4.2 provides a summary as well as a brief description of each type of document analysed.

Table 4.2 Summary of documents analysed and the sources of data

Document	Description	Data Source		
		University of Botswana	**Stellenbosch University**	**Makerere University**
Study guides	A study guide served primarily as a reference for students and faculty, to provide official information relating to a discipline clinical clerkship. It was available to all students and relevant faculty involved in that clerkship. It provided detailed information relating to the clinical clerkship, including the teaching schedules, duty rosters, projects and assignments that students were required to conduct and submit, assessment methods and resource materials.	Paeds: Overall clinical coordinator for the MBBS programme provided the study guide and logbook. O&G: HOD provided the study guide and logbook.	Paeds and O&G: Overall coordinator for the MB,ChB programme provided the study guides and logbooks for both disciplines.	Study guides were not available but relevant information was obtained from a copy of the MB,ChB curriculum, provided by the Paeds HOD, and also by the O&G HOD.
Logbooks	Logbooks were used to record or log the clinical procedures that students were required to observe and/or perform under supervision, whether in a clinical skills centre using mannequins for practice or in a clinical setting. Some procedures had a predetermined number and students recorded the details each time the procedure was performed.	(As above) Logbooks were incorporated into the study guides.	(As above) Logbooks were incorporated into the study guides.	Logbooks were used but not available

Abbreviations: HOD = Head of Department
O&G = Department of Obstetrics & Gynaecology
Paeds = Department of Paediatrics

Document	Description	Data Source		
		University of Botswana	**Stellenbosch University**	**Makerere University**
Assessment Schedules	Assessment schedules provided information regarding the frequency and timing of assessment during the clinical clerkship or rotation, and also the timing of exit-level assessments in the academic year.	Overall clinical coordinator for the MBBS programme	Examination Office, Faculty of Medicine & Health Sciences, Tygerberg Campus.	None available
Written Assessment: Questions	Written assessment formed one of two components of the exit-level assessments at all the study sites.	Overall clinical coordinator for the MBBS programme or Final Year Examination coordinator.	Paeds: Clinical clerkship coordinator.	Respective HODs for Paeds and O&G.
Written Assessment: Model answers/ Marking guides	Model answers or marking guides corresponding to the written assessment questions were created to assist examiners involved in marking the questions.	Overall clinical coordinator for the MBBS programme or Final Year Examination coordinator.	Paeds: Clinical clerkship coordinator	Respective HODs for Paeds and O&G.
Clinical Assessment: Questions	The other component of the exit-level assessments at the study sites, clinical assessment, took the form of Objective Structured Clinical Examinations (OSCEs) and Clinical Long Cases. OSCE questions were pre-determined and a printed version of the questions was available and displayed at each OSCE station. Clinical Long Cases involved the use of real patients, and questions asked were not pre-determined but varied depending on the clinical condition of the patient and the responses provided by the student during the examination time.	Obtained by field notes made while observing the actual clinical examinations.	Obtained by observations during the actual clinical examinations.	Obtained by observations during the actual clinical examinations.
Clinical Assessment: Checklists/ Marksheets	Structured pre-determined checklists or marksheets were created to assist examiners involved in marking the OSCE station. Marksheets were also used in clinical long cases to provide guidance to examiners	Not available	Paeds: Senior examiner (during clinical examination) O&G: Clinical clerkship coordinator	Paeds: Clinical clerkship coordinator. O&G: HOD

Study guides and logbooks, and assessment schedules, all of which were also available to the students, were requested from the overall programme coordinator or head of department. I was able to obtain a hard copy of the study guides and logbooks at Stellenbosch University for my own reference. At the University of Botswana, the overall MBChB programme coordinator provided me an electronic copy of the Department of Paediatrics study guide and logbook during my site visit. A printed copy of the study guide and logbook for the Department of Obstetrics & Gynaecology was provided for my perusal by the head of department, and, with their permission, I retyped my own electronic copy of the relevant information pertaining to the final year discipline clinical clerkship and assessment during a site visit before returning this document. I checked my electronic version against the original document for accuracy of copying. At Makerere University, the two respective heads of department each provided a printed copy of the MBChB curriculum, and similarly I retyped my own electronic copy of the relevant information before returning this document. This typed electronic version was checked against the original document for accuracy.

Other documents, such as the written assessment questions with the model answers and marking guides, and clinical assessment checklists or mark sheets that had more restricted access due to their confidential nature were also obtained on request after the scheduled examinations were over. These were used to provide information on the total assessment time, the range of methods used and their weighting (relating to the calculation of students' final assessment marks), the types and numbers of questions being asked and information expected of students (relating to course content). In addition, as University of Botswana written assessment papers were integrated, the clinical discipline involved was also recorded on the Excel spreadsheet. The clinical assessment questions were collated in the same way as with written assessment questions, using the same categories. Table 4.3 summarises the type of information sought during document analysis. Most of these documents were made available to me in electronic form and were stored in a secure locked facility to which I had exclusive access (see also under Ethical considerations, Section 4.7.3).

Table 4.3 Type of information sought during document analysis

Document	Information extracted
Study Guides Logbooks Assessment schedules	• Formative and/or summative assessment • Breakdown of components or types of assessment used • Weightings of individual components • Timing of assessments
Written assessment questions	• Clinical discipline • Type of assessment method(s) used • Number of questions • Question topics
Clinical assessment questions	• Type of assessment method(s) used • Number of questions • Question topics • Clinical discipline

In several instances at all three study sites, only the printed copy was available for me to peruse over a 24-hour period and then return. I created my own individual Excel files for each written and clinical assessment event in order to have a more structured and uniform way of summarising and presenting the information. Questions with the model answers and marking guides were retyped onto these Excel files, and rechecked against the original documents for accuracy of copying (See Appendix B).

During the process of data analysis, gaps in the data emerged. Similar experiences had been encountered in the preliminary study on assessment of clinical competence (Tan et al., 2016). Clarification was then sought with the clinical clerkship coordinators (faculty who were in charge of organising and coordinating individual clinical clerkships) or the respective head of department to verify the correctness of assessment-related information in the documents. This was done at the time they were interviewed with regard to other study objectives (See Section 4.4.3.1). So in addition, where appropriate, pertinent information obtained from my own field notes made after interviews with clinical clerkship coordinators and other study participants or during observations of clinical assessments was added to the information obtained from document analysis.

As far as possible, documents were accessed at the time of visits to study sites. However, on several occasions at the University of Botswana and Makerere University,

some documents (such as the written examination paper) were not immediately available during the site visits. Although assurances were provided at the time that the documents would subsequently be made available to me, this proved to be challenging and required perseverance in the follow-up via email correspondence, as obtaining a response to these requests for the relevant documents took between two to three months. This was a cause for concern, as the clerkship coordinators to whom the request was made might change their mind, with the level of interest falling and cooperation waning, as time went by. Nevertheless, this in-depth process of documenting and cataloguing was to provide for a rich source of data that would serve as foundation for the entire study.

4.4.2 Observations

Document analysis provided information on what exit-level assessment methods were stated as being used, as well as the nature of the actual assessment tools both in terms of structure and of content. The next step was to investigate the extent to which what was stated in the different study guides mirrored practice, and this could be done by observation. Observation "draws on the direct evidence of the eye to witness events at first hand" (Denscombe, 2010:196). This involved direct observation of what actually took place and taking notes of the conduct and process of the exit-level assessments. I conducted all of the observations and was guided in the process by a detailed observation schedule (See Appendix C) to assist with recording details of observations made while the exit-level assessments were taking place. This was formulated based on my own experiences as an assessor and clinical examination coordinator and refined following discussion with my supervisors who afforded comments before finalising the schedule. The purpose of the observation schedule was to minimise, and possibly eliminate, the variations that could arise from data based on observer perceptions of events and situations (Denscombe, 2010), and to provide a framework for observation which could be used consistently.

The designed observation guide, in the form of a checklist of items to assist me in monitoring and making a record of their absence or occurrence, had the following headings:

- Pre-examination preparation or briefings for examiners and/or students

- Station or room layout
- Details of clinical case or scenario
- Examiner: Comments and/or observations
- Student: Comments and/or observations
- Patient (if applicable): Comments and/or observations
- Equipment (if applicable): Comments and/or observations
- Post-examination discussions
- Observer notes

The focus of the observation was on the details of the clinical case or scenario. However, the additional notes provided background context of that observed assessment encounter. The guide allowed for the details of observations to be recorded as free text or sketches of room layouts at the time of actual observation (An example of a completed observation guide is provided in Appendix C2). Field notes were added later where required on reviewing the observation schedules, as it was not possible to record everything in full whilst observing.

The sampling strategy for the clinical examinations was by convenience sampling. All students and faculty involved were invited to participate in this study (The detailed description of this process is situated in Section 4.7.2 Informed Consent).

There were a number of technical challenges faced in the collection of data by observation, which included access to the exit-level assessment events, obtaining consent from various categories of participating individuals, space constraints of some examination settings, recording of observed data, and the possible influence of observation on the interactions between examiners and students. As mentioned earlier in Section 4.2.2, a key consideration for case studies as the choice of research design for the study was the need for study sites to be agreeable to being investigated. Seeking and obtaining institutional ethical approval, as well as identifying local investigators from Makerere University and the University of Botswana, helped provide support for the study, but these alone were insufficient measures. Faculty in Paediatrics at Stellenbosch University and the University of Botswana expressed some initial reservations when first approached about the study. These were addressed in a lengthy exchange of correspondence electronically with the respective Heads of Department, but arranging a face-to-face briefing for the whole department provided a further

opportunity to address any issues or concerns individual faculty had with the study. The approval of Heads of Department, whether explicit or implicit, played a major role in ensuring participation by department faculty, access to exit-level assessment documents and to sites for observation of clinical examinations. This issue of seeking prior approval from the management was a potential threat to participant autonomy and, by implication, to the principles of ethical research, and will be discussed further in Section 4.7.2.

4.4.3 Interviews

In a situation where simple, straightforward, factual information was required, a survey or questionnaire would suffice for data collection, and would be more cost-effective. However, interviews were used in this study "to explore more complex and subtle phenomena" in order to gain insights into the opinions, feelings emotions and experiences of the interviewees or participants (Denscombe, 2010:173-174). Interviews are a widely used method of data collection in qualitative research (Britten, 1995; DiCicco-Bloom and Crabtree, 2006). An interview is a conversation between two or more people in which the interviewer asks questions from the interviewee (Collins, 2017). Denscombe (2010:172), however, asserts that interviews are "more than just a conversation" as interviewees consent to participate, and the words spoken by the interviewee become publicly available. Although the agenda for the discussion is usually initiated by the interviewer, this may change depending on the responses from the interviewees. The purpose of the interview may set the tone of the interview and the degree of formality in the interaction, but it can still be conducted in the more relaxed manner of a conversation.

Denscombe (2010) describes various types of research interviews depending on the format of the questions and answers (structured, semi-structured, and unstructured interviews) and also by the number of study participants involved and the group dynamics (one-to-one, group and focus-group interviews). At all three study sites, semi-structured interviews were conducted. In this type of interview, a list of issues to be addressed is drawn up as a set of pre-determined open-ended questions to guide the interview (See Appendices B, C and D: Interview guides) with other questions arising from the responses of the interviewees (DiCicco-Bloom and Crabtree, 2006:315; Tavakol and Sandars, 2014b). The use of an interview guide provided consistency in

the way the questions were introduced and served as a prompt to ensure that questions would not inadvertently be omitted during the interview. It allowed flexibility in terms of the order of topics (Denscombe, 2010:175), and the responses could be open-ended. However, one possible limitation of having these pre-determined questions was that participants might only limit their replies to say what is expected of them. The challenge when conducting interviews would be to listen while encouraging interviewees to speak, and to be prepared to depart from the interview guide to follow and further explore the interviewee's responses (DiCicco-Bloom and Crabtree, 2006).

4.4.3.1 Interviews with faculty

All faculty members (including clinical clerkship coordinators and teaching faculty) from the Departments of Paediatrics & Child Health and Obstetrics & Gynaecology in the three medical schools who were involved in exit-level assessment, as examiners of the written and/or clinical examinations, or who contributed to the examination questions, were eligible to be interviewed. Interviews were conducted separately for faculty from the two disciplines at each study site, so as not to cause confusion in the discussion from faculty from these two disciplines.

The interviews focussed mainly on two of the study objectives (as displayed in Table 4.5), namely:

- To explore the reasons for the choices of exit-level assessment methods at the three selected medical schools
- To investigate what clinical competencies were considered appropriate for medical graduates from these medical schools in order to deliver selected health-related issues

and these informed the construction of the interview guide (See Appendices D and E for one-to-one interviews with clinical clerkship coordinators and group interviews with faculty respectively).

In addition to these two objectives, as mentioned earlier in Section 4.4.1, at the time clinical clerkship coordinators were interviewed, clarification was sought from them concerning gaps in information that emerged during document analysis.

Interviews with faculty comprised a mix of one-to-one and group interviews. As there was only a single participant in the one-to-one interviews, they were easier to arrange and relatively easy to control. The opinions and views expressed in the interview came from the single source. However, this would also mean the opinions and views were limited to that one person (Denscombe, 2010:176). This kind of interview was also easier to transcribe, as there would only be one voice to recognise. Group interviews, on the other hand, involved more than one person at a time. According to Denscombe, group interviews are conducted in a similar fashion to one-to-one interviews in that the interviewer is the focal point of the interaction (Denscombe, 2010:176-177). Responses are obtained from all members of the group, and more opinions and views could be gained for the same amount of time. Group interviews allow participants to listen to alternative points of view, and also to express support for certain views and to challenge the view with which they disagree (Denscombe, 2010:176-177).

Individual one-to-one interviews were the choice when interviewing clinical clerkship coordinators who were in overall charge of their respective clinical clerkships and who had the responsibility of ensuring the smooth delivery of the clinical clerkship content and assessment activities. They would most likely have an understanding of the clinical clerkship at a deeper level, so issues relevant to exit-level assessment practices in their respective departments could be probed more deeply at an individual and private level (Babbie and Mouton, 2001; Ritchie, 2003; DiCicco-Bloom and Crabtree, 2006).

The initial intention in the study was to conduct focus-group interviews. The focus groups, usually comprising between six to nine participants and a moderator (the researcher), are used to explore attitudes and perceptions, feelings and ideas about a specific topic (Denscombe, 2010:177). Interaction takes place within the group, and the moderator's role is to facilitate the group interaction, rather than lead. However, there were challenges in these resource-constrained settings to get sufficient numbers of clinicians to come together for focus-group interviews and discuss issues relating to exit-level assessment practices. Group interviews were therefore conducted in the study sites, involving groups numbering between four to six study participants. The main intention of the group interviews was to explore the perspectives of these individuals, how they talked about the issues and the effect of their opinions on other participants in the group (Ritchie, 2003). Where it was not possible to get groups of

four to six participants, they were interviewed in pairs or as a triad, and in two instances, on an individual basis to fit in with times convenient to their working schedules. Use was made of convenience sampling, although in the ideal situation, all faculty involved with exit-level assessment should have been invited to participate in the study. The interviews with two or three study participants still allowed them to hear each other's opinions and to compare with their own experiences and perspectives (Ritchie, 2003:37). There are advantages of both individual and group interviews. Benefit could be derived from the interaction or "joint reflection" in group interviews which would add to the richness of the data, and although participants in individual interviews would not have had this same benefit, they were still able to provide their own experiences and opinions. Table 4.4 at the end of this section provides a summary of the interviews conducted and the numbers of interview participants involved.

Data collection presented unique challenges in each context and an adaptive approach was utilised to access relevant respondents at each site. Different strategies proved effective in the different contexts, and this is described in detail below.

Heads of department in both disciplines were approached initially to inform them about the study and to invite departmental participation. Differing responses were obtained. It seemed that the agreement and support of respective heads of department was also crucial to ensuring the smooth conduct of the study at the three study sites. Faculty appeared more willing to participate if the study had the "approval" or acceptance of the head of department although this was not formally or officially required. It is plausible that this could have an impact on their participation. This issue will be discussed further in Section 4.7.2 on Informed Consent).

At Stellenbosch University the respective final (or sixth) year clinical clerkship coordinators in Obstetrics & Gynaecology and Paediatrics were initially invited to participate in the study by email. The final year coordinator in Obstetrics & Gynaecology, who seemed to be very supportive of this study, provided a list of names and contact details of examiners for me to contact by email. I also had opportunity to meet them in person at the time of observing the clinical assessments (OSCEs) in April 2014 and was able to invite them in person. Out of 23 examiners invited to participate in the study, five declined or replied they were not available for the study, and one expressed an interest but it was not possible to find a mutually convenient time due to

busy clinical schedules. There were no responses from the remaining examiners. In January 2015, a second attempt was made to contact examiners who were conducting an OSCE for postgraduate Obstetrics & Gynaecology trainees, with the help of a junior department coordinator. They were approached in person and invited, with the intention to conduct the interview on the same day as that OSCE. Out of the eight examiners listed for the OSCE, one could not be contacted; four declined to participate, two agreed and were interviewed on that same day. The remaining examiner also agreed but scheduled a meeting on another day, and attended the interview together with the junior department coordinator. In summary, two interviews were conducted in Obstetrics & Gynaecology each with a pair of study participants.

In the Department of Paediatrics at Stellenbosch University, the Head of Department expressed reservations about the study and advised that a presentation should be made regarding the study prior to the clinical examinations at a regular weekly department meeting attended by faculty. This was carried out in April 2014 before observations of actual exit-level assessments (clinical long cases) were scheduled. The presentation was followed up with an email invitation to the department faculty involved as examiners. Out of 20 examiners invited to participate in the study, five responded and agreed to participate. However, one was unable to attend the designated interview date due to her clinical schedule and no alternative dates were available. Over January and February 2015, a further attempt was made to contact examiners for interviews. This led to three further interviews with one, two and four study participants respectively. Overall, a total of four interviews were conducted with paediatric examiners.

At Makerere University, attempts prior to the site visit to set up an interview with the final (or fifth) year clinical clerkship coordinator in Obstetrics & Gynaecology failed as the Head of Department did not provide the necessary contact details. The arrangements to interview the clinical clerkship coordinator and examiners were therefore made at the time of the site visit in May 2014, after meeting with the Head of Department. This took place on the same day as the written assessment paper was being conducted, and as the clinical clerkship coordinator was the invigilator, it meant his interview could take place after the written assessment paper was over. One faculty member with whom I had prior acquaintance assisted in approaching department colleagues who were involved in the exit-level assessments to participate in the study.

Two group interviews with four and five study participants respectively were conducted on the same day.

The Head of the Department of Paediatrics at Makerere University was publicly very supportive of the study and had arranged for the department faculty to be available for group interviews. These were also conducted during the site visit in May 2014 on two consecutive days, one with six study participants and the other with four. An individual interview was conducted with the paediatric clinical assessment coordinator to provide a further perspective on exit-level assessment; the joint clinical clerkship coordinators in Paediatrics having already attended one of the two group interviews for examiners.

At University of Botswana, the Head of the Department of Obstetrics & Gynaecology ~~was~~ expressed full support for the study, and provided the email addresses of individual department faculty to enable me to contact them myself as he did not want to influence their decision to participate in the study. Three names with contact details were provided. One responded that he was no longer working with University of Botswana, a second declined to participate, and the third agreed to participate in the study. There were no other faculty available as replacements. The Head of Department of Paediatrics also agreed to support the study, but circulated my invitation to participate to the department in order to elicit the views of individual faculty in this regard. The department consensus was to support the study, and the Head of Department then provided the name and contact details of the final year clinical clerkship coordinator as he would be responsible for organising interviews with faculty who were involved in exit-level assessment as examiners. One group interview with four faculty was arranged. These interviews took place in June 2015 during the site visit.

At all three study sites, an electronic copy of the participant information sheet and consent form was provided for reference at the time of initial email invitations. A convenient date and time for each interview was determined, and a suitable venue arranged. At the start of the actual interview, the study was discussed and the participants given the opportunity to have their questions addressed. Each study participant was provided a printed copy of the participant information sheet and consent form to sign.

A total of 23 interviews with 68 participants (examiners and interns) were conducted at the three study sites. They varied in length, the shortest being 16 minutes and 30

seconds, and the longest, 72 minutes and 26 seconds. The average duration was approximately 50 minutes. A summary of the interviews is provided in Table 4.4 (The findings will be discussed later in Chapter 5). I conducted all the interviews myself and audio-recorded them with the aid of a digital-voice recorder. Copies of the recordings were made electronically on my password-protected computer for back-up. Although the numbers participating in the interviews appear small, the views expressed by the interview participants could be considered to be representative of their respective disciplines, since the overall number of faculty in the disciplines was not large. The human resource constraints that are a feature in the Sub-Saharan African context and have already been highlighted in Chapter 3 would account for these small numbers of faculty.

Table 4.4 Summary of interviews conducted in relation to specific study objectives

Study Objectives	Data Source								
	Clinical clerkship coordinators			Faculty involved in assessment			Interns (Graduates of the programme)		
	University of Botswana	Stellenbosch University	Makerere University	University of Botswana	Stellenbosch University	Makerere University	University of Botswana	Stellenbosch University	Makerere University
• To investigate what ELA methods were stated as being used in resource constrained settings • To explore the reasons as to why these methods were chosen • To investigate what clinical competencies were considered appropriate for medical graduates from these medical schools in order to address selected health-related issues	O&G: HOD (1) Paeds: Final year coordinator (1)	O&G: Final year coordinator (1) Paeds: Final year coordinator (1)	O&G: Final year coordinator (1) Paeds: Clinical exam coordinator (1)	O&G: (1) Paeds: (4)	O&G: (2) (2) Paeds: (4) (1) (2) (4)	O&G: (4) (5) Paeds: (6) (4)			
• To investigate what ELA methods were actually being used in resource constrained settings								(5) (5)	(5) (3) (5)

Abbreviations: ELA = Exit-level Assessment
HOD = Head of Department
MDGs = Millennium Development Goals
O&G = Department of Obstetrics & Gynaecology
Paeds = Department of Paediatrics

Notes: Numbers in brackets refer to number of interviewees/study participants in an interview

Digital-audio recordings of the interviews were sent directly for transcription in the form of electronic password protected files via email to a professional transcription service provider (transcriber) who was not attached to any of the medical schools in the study. In order to assist the transcriber with recognition of the voices of participants, the transcriber was told the number of participants in the interview and the order in which they first spoke, with each participant assigned initials. I was assured that all digital-audio recordings and correspondence were handled by the transcriber in the strictest confidence. The audio recordings were stored on her personal computer which was password protected, and to which only she had access. As each audio recording was transcribed, it was emailed to me for checking that it accurately matched the original digital-audio recording (A sample of an interview transcript is provided for reference in Appendix F). On completion of all the transcribing to be carried out for the study and after I had informed her that I did not have any further queries regarding any of the transcriptions, her electronic copies of the transcriptions as well as the digital-audio recordings kept on her computer were all deleted.

A separate electronic file was created for each interview, and assigned a unique reference number for cataloguing and indexing purposes. After the interviews had been transcribed, I performed quality control on the data to ensure that the information was captured as accurately as possible. Measures taken to ensure trustworthiness are discussed later in Section 4.8.3. After the necessary corrections had been made and the transcriptions were checked a final time, initials of participants were replaced by a unique code number that was assigned to each participant and the transcript files were ready for data analysis (described in Section 4.6 to follow).

As with observation of exit-level assessments, some technical challenges were encountered during the conduct of interviews. These included having, on occasion, to endure noisy venues with interruptions such as some participants arriving late or having to leave before the end of the session, cellular telephones ringing and also electricity power cuts. More desirable and conducive settings for conducting qualitative research interviews were not always available and it was necessary to manage with these less than ideal situations as best as one could. Having adequate time to conduct and complete the interviews also posed a challenge on occasion (as with the Obstetrics & Gynaecology clinical clerkship coordinator at Makerere University) although in most

instances, the time was sufficient. Some examiners had heavily accented speech and spoke very fast, making it difficult to pick up all that they said during the interview. As mentioned earlier in this section, following the subsequent task of transcribing, I paid particular attention to checking the transcriptions with the digital-audio recordings in order to ensure as much accuracy as possible before sending these back to interview participants to further check for correctness.

4.4.3.2 Interviews with interns

Interns who graduated within the last two years (in 2014 and 2015) from Stellenbosch University and Makerere University were eligible to be participants in this study. The first cohort of students from University of Botswana graduated in June 2014 and was not included in the study as applications for ethical approval from the institutional review boards were made prior to this date. While other young doctors who had completed their internship and progressed to performing their community service would have gained some working clinical experience and may have provided a valuable different viewpoint from those of interns, the choice of interns in this study was for two reasons: (i) exit-level assessment experiences would be more recent and fresh in their memory, and (ii) the main focus of undergraduate medical-training programmes was to prepare the students for internship on graduation.

Interviews with interns focussed on these two study objectives:

- To investigate what exit-level assessments were actually being used at two of the three selected medical schools (which would also be a means of corroborating information obtained from document analysis and from observation of the exit-level assessments)
- To investigate what clinical competencies were considered appropriate for medical graduates from these medical schools in order to deliver selected health-related issues

Interns were asked about their recall of the exit-level assessment and for their views on the appropriateness of the exit-level assessments for the context they now worked in.

A convenience sampling approach was initially adopted to recruit interns. As interns were usually assigned to selected hospitals nationwide for the duration of their internship under supervision, for practical reasons in this study, interns who were

graduates of the two medical schools (Stellenbosch University and Makerere University) and who were based at the main teaching hospital affiliated to the respective medical school were invited to participate in the study. Stellenbosch University graduates who were undertaking their internship at Tygerberg Hospital in Western Cape Province were identified through the University Alumni Office and invited to participate in the study by email. The email invitation was adapted from information already in the participant information leaflet, providing background to the purpose of the study and what the interview process would entail. However, these invitations drew just one response out of 13 invitations, in spite of several attempts. That intern who responded was contacted and asked to help recruit his intern colleagues for the study, in a form of "snowball sampling" ("snowball" referring to "the process of accumulation as each located subject suggests other subjects" (Babbie and Mouton, 2001:167). A group interview with five interns was conducted in December 2014 using an interview guide (See Appendix G). At that time, this group of interns had completed their rotations in Paediatrics but not in Obstetrics & Gynaecology. In order to arrange for a further group interview with interns who had undergone the Obstetrics & Gynaecology rotation in their internship, the intern coordinator in the department, who was contacted and asked for help, arranged for the current interns to participate in the study. This second group interview with five interns was conducted in June 2015.

At Makerere University, a similar snowball sampling technique was used to locate interns undertaking their internship at Mulago Hospital and other hospitals in Kampala, starting with one intern who was known to the Head of the Department of Paediatrics and Child Health. The intern contact arranged for colleagues to participate in interviews at Mulago Hospital (two groups each with five participants) and one triad at Nsambya Hospital also situated in Kampala. These interviews all took place during the site visit to Makerere University in May 2014.

Thus to sum up, this section has described the various sources of data and the methods used for data collection. These are all summarised in relation to the study objectives in the following Table 4.5.

Table 4.5 Matrix of methods of data collection used per research study objective

<table>
<tr><td></td><td colspan="4">DATA SOURCES</td></tr>
<tr><td>STUDY OBJECTIVES</td><td>Document analysis</td><td>Observations</td><td colspan="2">Interviews</td></tr>
<tr><td>• To investigate what ELA methods were stated as being used in resource constrained settings</td><td>Documents relevant to ELAs
(See also Table 4.3)</td><td></td><td>Interviews (individual) with clinical clerkship coordinators</td><td></td></tr>
<tr><td>• To investigate what ELA methods were actually being used in resource constrained settings</td><td></td><td>Observation of ELAs</td><td></td><td>Interviews (in groups) with interns</td></tr>
<tr><td>• To explore the reasons as to why these methods were chosen</td><td></td><td></td><td rowspan="2">Interviews (individual or in groups) with clinical clerkship coordinators and faculty involved in ELAs</td><td></td></tr>
<tr><td>• To investigate what clinical competencies were considered appropriate for medical graduates from these medical schools in order to address selected health-related issues</td><td></td><td></td><td>Interviews (in groups) with interns</td></tr>
<tr><td colspan="5"></td></tr>
<tr><td>Overall research question:
To investigate the validity of ELAs in determining the clinical competence of Sub-Saharan African medical school graduates to address selected health-related issues</td><td colspan="4">Analysis of data generated by the previous phases of the study (as listed above)</td></tr>
</table>

[Abbreviation: ELA = Exit-level assessment]

4.5 Data Collection

4.5.1 Timing of data collection and timeframe of study

Visits to the three study sites were timed to coincide with the scheduled exit-level assessments so that observations could be made. Data collection took place between April 2014 and June 2015. At Stellenbosch University, this occurred over a two-week period in April 2014. Exit-level assessments at Makerere University took place over a 10-day period at the end of May/beginning of June 2014. The examinations at University of Botswana took place in July 2014, but as ethical approval for the study had not been granted in time from the University, the visit to University of Botswana was deferred until the first week of December 2014 when the supplementary examinations were being conducted for three students who had been referred from the July examinations. This visit provided an opportunity to meet the heads of department and clinical clerkship coordinators in Paediatrics and Obstetrics & Gynaecology in person as prior attempts to contact them electronically had been unsuccessful. A second site visit was made to University of Botswana over a 10-day period in June 2015 during the following round of exit-level assessment, as the main exit-level assessments were conducted at this time in the academic calendar.

Table 4.6 Timing of data collection

	University of Botswana	Stellenbosch University	Makerere University
Document analysis	June 2015	April 2014	May 2014
Observation of ELAs	June 2015	April 2014	June 2014
Interviews	Dec 2015, June 2015	Between April 2014 – June 2015	Between May – June 2014

Abbreviations: ELAs = Exit-level Assessments

For pragmatic reasons, document analysis and the conduct of interviews needed to be performed in close relation to the observations, in particular, at Makerere University

and University of Botswana in order to optimise the duration of site visits and costs involved whereas the timing of data collection at Stellenbosch University did not require the same level of urgency since I was resident in South Africa for the greater part of the study. Table 4.6 summarises the timing of data collection.

In terms of the order in which the different methods of data collection were conducted, document analysis was done first in order to obtain an overview of assessment methods stated as used. Observation of exit-level assessment practices was the next logical step to verify the match between what was stated and actual practice. The third step was then to conduct the interviews, to clarify information gaps in documents relating to exit-level assessment and to verify the correctness of the information provided in study guides (where relevant with clinical clerkship coordinators), and to further explore issues around the exit-level assessments.

4.6 Data Analysis

Large amounts of data are often collected in qualitative studies, which necessitate ways to make sense of it all (Pope et al., 2000; Tavakol and Sandars, 2014b). In this study, data in the form of text was generated. Many approaches are available for analysing text data and focussing on interpreting the meaning. Content analysis is one such method used widely in health professions education (Hsieh and Shannon, 2005; Elo and Kyngäs, 2008; Wang et al., 2014; Tavakol and Sandars, 2014b), either with an inductive or deductive approach. An inductive approach provides a "bottom-up method of analysis" (Tavakol and Sandars, 2014a:748). In the case of this study, inductive analysis was used with the data obtained from interviews. A deductive approach, on the other hand, "begins with formulating a research hypothesis about the phenomenon of interest" (Tavakol and Sandars, 2014a:748), and was applied to data relating to assessment content. These approaches will be described further shortly, but first, prior to analysis the data needs to be prepared.

Hsieh and Shannon (2005:1278-1279) refer to data analysis as "qualitative content analysis" and define it as "a research method for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" and propose that it is suitable for use in studies (such as this one) which aim to understand a phenomenon, when "existing theory or research

literature on a phenomenon is limited". The data analysis process essentially involves a number of steps (Hsieh and Shannon, 2005; Elo and Kyngäs, 2008; Wang et al., 2014; Tavakol and Sandars, 2014b). The steps used by Creswell (2014) are described here as applied to this study, on the left-hand side with reference to Figure 4.1. Elo and Kyngäs' model (2008) describing three main phases in the data analysis process: preparation, organising and reporting, are also shown on the right-hand side in Figure 4.1 to provide another perspective which has broad overlap with Creswell's model and can be used to complement each other.

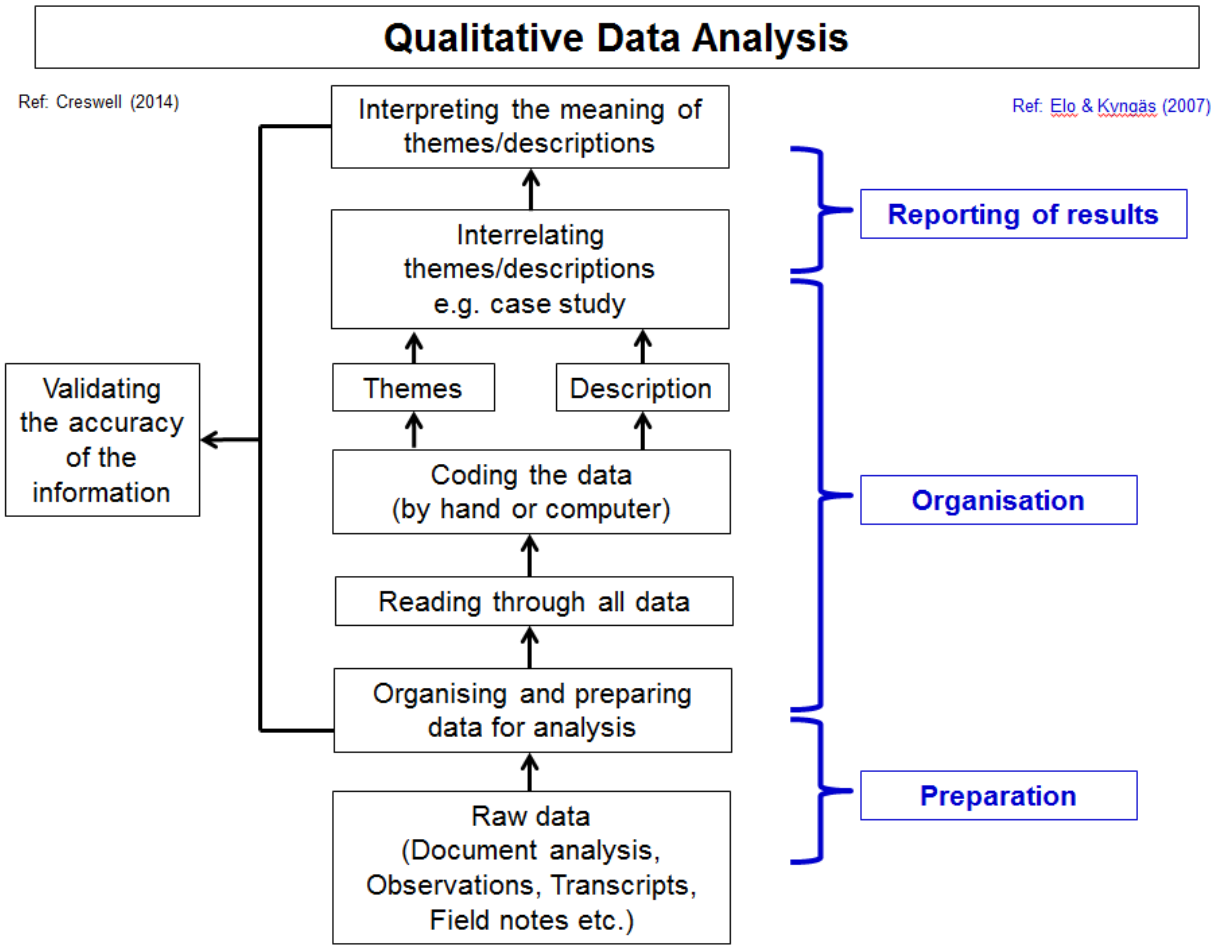


Figure 4.1 Steps in validating the accuracy of the findings

References: Elo and Kyngas (2008)
Creswell (2014)

Starting at the lower end of Figure 4.1 and progressing upwards through the steps in ascending order, data gathered from document analysis, observations, interviews and

field notes first need to be organised and prepared for data analysis, and this applied to all forms of data collected in this study. Given the multiple data sources and the overall complexity of the study, a combination of both inductive and deductive approaches to data analysis were used.

4.6.1 Inductive analysis of interview data

Following the transcription of audio recordings of the interviews into text (already described in Section 4.4.3), the next step is for the data to be organised into "meaningful units" (Tavakol and Sandars, 2014b). This includes careful reading and re-reading of the text of transcriptions to look for text with similar meaning. Using a coding process, the text is organised into categories – "patterns or themes that are directly expressed in the text or derived from them through analysis" (Hsieh and Shannon, 2005:1285). As the categories are derived from the data and not based on previous knowledge or pre-determined, this forms the basis of inductive content analysis (Elo and Kyngäs, 2008:109). A coding scheme is drawn up, based on the data that has been organised into categories, and this is used to guide researchers involved in the coding process to make decisions in the analysis of content (Hsieh and Shannon, 2005:1285). Where categories share similar patterns or descriptions, themes can be identified, contributing to results or findings of the study. These themes then need to be interpreted, in order to make overall sense of the data.

In the case of this study, all coding was done by me. The first transcript was picked at random and used to create the initial coding scheme, which was done by hand. With subsequent transcripts, the codes were then added to and amended. Cross-checking of codes was undertaken by one study supervisor who reviewed two transcripts and the proposed coding that I had carried out. In this way, a code book (or code list) was created, providing the compilation of the codes, their content descriptions, and a brief example for reference (Saldaña, 2013:24-25). The final list of codes is provided for reference in Appendix H.

With large amounts of data, the whole process can take considerable time and effort. In this study, version 6.2.28 of ATLAS.ti™ (Muhr, 1997), a computer-assisted qualitative data analysis software (CAQDAS) programme was used to facilitate this process of qualitative data analysis, enabling the complex organisation and retrieval of data (Pope et al., 2000). Although CAQDAS has considerably reduced the time spent coding the

data, it cannot be regarded as a replacement for the whole data analysis process. The researcher still has overall control and needs to decide the codes, categories and themes through careful analysis of the transcriptions and textual data (Tavakol and Sandars, 2014b).

4.6.2 Deductive analysis of data from documents and observations

One other aspect of data analysis related to the content of exit-level assessment in the form of the actual assessment questions, collected from document analysis and observation. The documents relating to exit-level assessment were mapped using an Excel spreadsheet to generate an overall view of the exit-level assessment practices taking place in each of the three medical schools. The criteria used to guide this mapping were based on those used in the preliminary study on the assessment of clinical competence in one Sub-Saharan African medical school (Tan et al., 2016). As explained earlier in Section 4.4.2, an observation schedule was used to collect data during observation of the actual clinical assessments. This data was supplemented by field notes written after the assessments, addressing queries which arose during the observations. The topics or themes of individual examination questions in each assessment method were tabulated as the first step in the analytical process. Two analytical tools were then used for deductive analysis of data from the documents and observations made.

Miller's (1990) 'pyramid' was used as an analytical tool in relation to assessment practices and to the appropriateness of the assessment content for exit-level. The data analysis will be referred to again in Sections 5.2.3, and 5.3.1, but briefly here, examination questions were each assigned to a tier of the 'pyramid' (the "Knows", "Knows How", "Shows How" and "Does" tiers) with reference to the descriptions Miller used for each tier (Refer back to Figure 2.1). Examination questions that brought out the recall of factual knowledge in students were assigned to the "Knows" tier. Questions that made students apply their prior knowledge in order to solve or manage a clinical problem were assigned to the "Knows How" tier. This could include clinical problems or scenarios in which information was provided (such as clinical information relating the symptoms or physical findings in a patient, laboratory or other diagnostic investigations, treatment therapies, and so forth). The student would be expected to analyse and interpret all the information provided and formulate a diagnosis and/or

plan further management for the patient. For examination questions that were assigned to the "Shows How" tier of the 'pyramid', the student not only possessed the required knowledge and was able to apply it to the relevant context, but was also expected to demonstrate pertinent skills for the management of the patient. Skills here could include communication skills (for example, taking a history, providing an explanation on a clinically-related topic to a patient, dealing with patients in challenging situations), physical examination skills, and also performing certain clinical procedures. These skills could be demonstrated to an examiner with the aid of a real patient, a simulated patient, or on a mannequin or teaching aid. The fourth tier, "Does" referred to actions or tasks that the student would undertake in clinical practice as a routine and not perform only in a more artificial assessment context.

Since it was published, there do not appear to be any studies that have actually made use of Miller's 'pyramid' as an analytical tool in a similar way to how it has been used in this study, and the majority of the numerous citations made in other studies are with reference to the utility of Miller's pyramidal framework for the assessment of clinical competence (van der Vleuten, 2000; Wass et al., 2001c; Crossley et al., 2002; Norcini, 2003c; Davis, 2003; Schuwirth and Van der Vleuten, 2003b; van der Vleuten and Schuwirth, 2005; Hodges, 2006; Carraccio and Burke, 2010). A smaller number of studies have expanded on the original 'pyramid' framework (Rethans et al., 2002; Cruess et al., 2016) as a concept. In this study I made use of Miller's 'pyramid' to investigate exit-level assessment and the clinical competence of medical graduates in a resource-constrained context in a novel way as an analytical tool, and this also has potential application in other settings. Its application in the analysis of assessment questions items is irrespective of the assessment method used.

Deductive coding was then applied here in this process to test Miller's model of the assessment of clinical competence by providing some measure of the validity of the exit-level assessments in determining clinical competence. Elo and Kyngäs used deductive content analysis "when the structure of analysis is operationalized on the basis of previous knowledge" (2008:107). Thus the analysis of the data collected for this research did not only comprise the linear/vertical process as depicted in Figure 2.1, but was further subjected to a horizontal analysis against a predetermined framework.

Using Miller's 'pyramid' as an analytical tool for this study was a useful exercise and provided an opportunity to really confirm its applicability. However, although it seemed easy to use in principle, it was actually very challenging because categorisation of examination questions required judgements (based on a combination of educational knowledge and clinical experience) to decide to which tier of the pyramid the questions should be assigned. These judgements or interpretations of the questions could be open to subjectivity, and this would need to be taken into consideration. Findings reported in Chapter 5 will provide examples to support these statements. As with inductive analysis of the interview data, this process of coding by deductive analysis was reviewed by one study supervisor who is an expert in the field of assessment in Health Professions Education.

The second analytical tool was Disability Adjusted Life Years (DALYs), which was used in relation to the content validity of exit-level assessment questions (a detailed description of the data analysis will be provided later in Section 5.3.2). As described in Section 3.1, at the start of this study selected health-related Millennium Development Goals (MDGs) were considered as appropriate measures to investigate the validity of exit-level assessments. MDGs 4 and 5 corresponded to the clinical disciplines of Paediatrics and Obstetrics & Gynaecology respectively and contributed to the decision to focus on these two clinical disciplines for the study. However, following initial data collection especially during interviews with examiners, it became evident that the MDGs were more appropriate at the level of informing overall global health policies but were not specific enough to inform at the level of curriculum and assessment. Global Burden of Disease (GBD) studies were then considered, since these related to health-care issues that could be measured by World Health Organisation (WHO)-defined regions. The indicators used in the analysis of results obtained in the GBD studies included disability-adjusted life years (DALYs), which provided a more detailed representation of the burden of disease and, therefore by way of implication, the health-care issues that would be relevant to that particular WHO region. On further review of this measure, DALYs were chosen as the substitute for MDGs.

Disability-Adjusted Life Years (DALYs) measure overall disease burden and population health and are expressed as "the sum of the Years of Life Lost (YLL) due to premature mortality in the population and the Years Lost due to Disability (YLD) for people living

with the health condition or its consequences" (World Health Organisation, 2015). The information gathered from countries worldwide can be utilised as indicators of regional health issues. Information obtained from DALYs and studies on global burden of disease have mainly been used to help shape global health policies. Similarly to Miller's 'pyramid', there does not appear to be any literature relating to the use of DALYs as a measure or reference point for clinical competence at exit-level of a medical training programme, and this study has made an attempt to do so.

The choice of DALYs as an analytical tool for the content of assessment is open to question. Even though DALYs have proven to have greater utility than MDGs for this purpose, this study has revealed challenges in matching assessment content to DALYs. It is by no means a perfect match, but in the absence of alternative satisfactory tools for analysis of assessment content, it could be considered a starting point for further research. Its strength lies in the fact that the causes of DALYs provide some contextual measure of disease conditions for the respective region using reference data from the World Health Organisation, thereby taking advantage of existing data without incurring added costs of collection and analysis. In hindsight, DALYs would still provide a better match to burden of disease for specific WHO world regions.

4.7 Ethical Considerations

Objectivity and integrity are two guiding principles in the conduct of scientific research (Babbie and Mouton, 2001). As ethical issues arise in any research study, prior consideration needs to be given to how these are addressed in the study (Lewis, 2003; Tavakol and Sandars, 2014b). Most ethical issues focus on the need to protect study participants, but there is also the need to take into account wider moral and ethical implications of the research being conducted (Babbie and Mouton, 2001; Lewis, 2003; Denscombe, 2010; Plowright, 2011). In the section that follows I discuss how these issues were addressed.

4.7.1 Institutional Review

There is a need for ethical oversight to ensure there is independent review of the study and for public accountability. Ethical approval to conduct this study was obtained from the Institutional Review Boards of the all three participating medical schools in the following order:

(1) Stellenbosch University Faculty of Medicine & Health Sciences Health Research Ethics Committee (Reference Number: S14/02/045, granted 20th March 2014) (See Appendix I),

(2) Makerere University College of Health Sciences Research Ethics Committee (Reference Number: #REC REF 2014-057, granted 15th April 2014) (Appendix J), and

(3) University of Botswana Office of Research and Development, (Reference Number: UBR/RES/ETHI/110, granted 3rd October 2014) (Appendix K).

At Makerere University, the Head of Department of Paediatrics and Child Health was invited to be the local co-investigator in the study, and helped to facilitate the submission of the research proposal to the Makerere University College of Health Sciences Research Ethics Committee for approval. Approval was also sought from and granted by the Uganda National Council of Science and Technology (a compulsory requirement for all research conducted in Uganda) (See Appendix L).

At the University of Botswana, the Associate Programme Director at the School of Medicine was invited to be the local co-investigator in the study, and helped to facilitate the submission of the research proposal to the Office of Research and Development, University of Botswana. This process took considerably longer than at the other two medical schools, and nearly six months elapsed between the submission of the request and the granting of approval.

Research was not commenced at the respective sites until formal approval was granted. Ongoing approval was obtained from Stellenbosch University Faculty of Medicine & Health Sciences Health Research Ethics Committee annually throughout the research activity, although approval from Makerere University and the University of Botswana was only required for the duration of one year (which was sufficient time for data collection to be completed at those two sites).

4.7.2 Informed Consent

Participation in research should be voluntary, with mechanisms in place to ensure this principle is upheld. The participant information leaflet and consent form, constituting the informed consent for this study, was based on the template provided by the Stellenbosch University Faculty of Medicine & Health Sciences Health Research Ethics Committee and was adapted to suit the context of each individual study site. It

addressed ethical principles centred on autonomy (respect for the study participants), beneficence (to "do good", or the moral obligation to act for the benefit of study participants), non-maleficence (to "do no harm" to study participants) and justice (fairness in distribution of any benefits or disadvantages of the study) (Tavakol and Sandars, 2014b).

Culturally and linguistically appropriate versions of the written informed consent forms were made available to study participants. For reference, the English language versions of the participant information leaflets and consent forms use are provided in Appendices M1 (for interview participants), N1 (for assessors), O1 (for students) and P1 (for patients), with the corresponding Afrikaans language versions in M2, N2, O2 and P2 respectively. Interview participants, assessors and students at Stellenbosch University were provided a choice of participant information leaflets and consent forms in English and Afrikaans. In addition to English, the documents used for patients were translated into Afrikaans for use at Stellenbosch University (Appendix P2), Luganda for use at Makerere University (Appendix P3), and Setswana for use at University of Botswana (Appendix P4).

Informed consent for study participants was obtained after the purpose of the study was explained to them verbally in addition to their being provided with the participation information leaflets. In particular, they were assured of full confidentiality and their right to withdraw from the study at any time, although none did so. Study participants who were interviewed were also informed that they would each be allocated a unique identification code number in order to protect their identity if any quotations arising from the interview transcripts were used.

Consent was obtained from students (for observation of the clinical examinations) at a different time from examiners and patients involved in the clinical examinations for pragmatic reasons. As it was considered too stressful and inappropriate to obtain consent on the morning just prior to the start of the examinations students were approached on the day before their clinical examinations, such as at the end of the last written paper before they left the examination venue. The class was addressed collectively and a brief explanation of the study was provided in addition to the participant information sheet and the consent form. A small number of students at Stellenbosch University chose not to agree to being observed during their examinations.

A note was made of their names so that they could later be identified on the morning of the Obstetrics & Gynaecology OSCE examination, and a small red sticker dot placed on the posterior aspect of their shoulder. This was not visible to examiners, but it provided an easy visual marker to assist me in withdrawing myself from observing the station that they were at. This process was not required for the Paediatric long clinical cases as none of the students who did not consent to participating in the study were encountered for the duration of my observation in that examination.

Consent of examiners and patients was obtained on the morning of the examinations, as I did not have prior details of participating individuals and therefore could not undertake this task earlier. The time constraint made it very challenging to get this all done prior to the start of an examination while giving due consideration to ensuring that participants under already stressful conditions were genuinely willing to take part in a research study.

There were no objections from any examiners and patients to being observed. However, I was mindful of the possibility that patients may have felt a sense of obligation to participate in the examinations as it may have provided an opportunity for them to receive expedited health-care (for example, being on the waiting list for a scheduled surgical procedure), and as they were already participating in the clinical examinations, they could be more likely to agree to be part of the research study. I made a point of reminding all study participants verbally (even after they had signed the informed consent form) that it was each participant's free choice to take part, and they had the right to change their mind and withdraw from the study at any time.

Examiners were, on the whole, helpful and accommodating especially in some of the examination settings where space was very limited. For example, some of the multi-purpose teaching rooms that were used at the University of Botswana for the conduct of the OSCEs were small, with dimensions measuring approximately 2.5 metres by 3 metres and had only sufficient space for an examination couch, a washbasin, and at most three chairs. It was difficult at times to find a suitable corner from which to observe, and on occasion it would involve standing right by the door. As an observer in a non-participant capacity (Babbie and Mouton, 2001:293), I strived to remain unobtrusive as far as possible, being always mindful of the influence my presence could have on the examiner-student interaction. Steps taken to achieve this included standing

behind the student and/or out of the direct gaze of either examiner or student, keeping as quiet as possible, avoiding any sudden or obvious movements, and adopting a technique of observing "indirectly" while averting my gaze.

As indicated earlier in Section 4.4, in the course of recruiting study participants and obtaining informed consent, there were several instances when individual autonomy of participants was potentially jeopardised. Tacit approval of the respective head of department for the study to be conducted appeared crucial, even if clinical clerkship coordinators or individual faculty were agreeable to participate in the study. There is also the possibility that they may have felt pressurised to participate in the study. However, during the interviews, study participants appeared to be genuinely willing to participate, as reflected in some instances by lively discussions that took place, and also by the information that was offered.

4.7.3 Confidentiality and storage of data

Anonymity and confidentiality are important in order to protect the interests of the study participants as well as their identity, and as part of respect for these individuals. All study participants who were observed at the time of the exit-level assessments were assured of anonymity and that any information gathered would be kept strictly confidential at all times. It was stressed to the observed participants that observations focussed on the process and content of the assessment, and not on the individuals involved. All exit-level documents and observation guides (in hard copy format) were stored in a secure locked facility to which I had exclusive access.

All study participants in the interviews (clinical clerkship coordinators, examiners and interns) were assured that any information gathered would be kept strictly confidential at all times. The date and time of the interview were recorded, but all personal identification was removed from the records. All data collected on electronic storage devices and in hard-copy format (including original audio recordings of interviews, field notes and paper documents, and the lists of interviewees which was kept separately from the interview data) were stored in a secure locked facility and analysis of the data was stored on a password-protected computer to which nobody else had access. Audio recordings and transcriptions of all interviews will be destroyed at the end of the research study, after the research findings have been published in peer-reviewed scientific journals.

4.7.4 Honesty and integrity

Just as a practising doctor, I have strived to adhere to the professional code of conduct for physicians at all times, I am bound by the professional code of ethics in research. As a researcher, I should operate with honesty and integrity especially with regard to the study. This includes dealing honestly with the study participants or subjects and not deceiving them, such as by lying about the purposes of the research. Honesty and integrity also apply in terms of accurate reporting and analysis of findings. Any negative findings should equally be included and not omitted. There should be no fabrication or falsification of any data.

4.7.5 Other study ethical issues

4.7.5.1 Funding and conflicts of interest

Funding and any conflicts of interest which could be considered as potential bias should be declared. I have no conflict of interest here, especially as I do not have any professional affiliations to the medical schools that were investigated in this study. Funding was received for this study from two sources: (1) the U.S. President's Emergency Plan for AIDS relief (PEPFAR) through the Health Systems and Resources Administration (HRSA) and managed by the Stellenbosch University Rural Medical Education Partnership Initiative (SURMEPI) and (2) from the Foundation for the Advancement of International Medical Education and Research (FAIMER), a U.S.-based non-profit foundation.

4.7.5.2 Anonymization

Anonymization of the three medical schools for this thesis has been impossible and, therefore, pointless since it would be easy to deduce from the information or descriptions provided. An example to illustrate this point is that the University of Botswana is the only medical school in Botswana.

4.7.6 Research Ethics and Society

There is a general obligation to conduct research in a socially responsive and responsible manner, and the guiding principle in the relationship between conducting research and the rest of society is accountability (Babbie and Mouton, 2001). As this study has received funding from two sources, there is also a need to be answerable to

the sponsors. Research findings should be made available to the public domain and openly disseminated, with consideration from the perspectives of educational value and scientific validity.

4.7.6.1 Educational value

The findings of this research will be presented to the MBChB programme committees of the respective medical schools involved, and also in the form of a workshop organised at the end of the study for all faculty from the Departments of Obstetrics & Gynaecology and Paediatrics & Child Health in the respective medical schools. The findings will further be presented at suitable conferences at national and international levels and will be submitted for peer-reviewed publication in suitable scientific journals.

4.7.6.2 Scientific validity

Transparency in assessment is a challenge in higher education and particularly so in medical education where a range of assessment methods are employed to assess competence. It is hoped this research would add to the body of knowledge, and provide insights into exit-level assessment for resource-constrained settings. In providing evidence of actual assessment practices in relation to health-related outcomes required to respond to the needs of Sub-Saharan Africa, this would help inform assessment practices going forward, including the decision-making processes on the clinical competence of medical graduates, and could potentially impact on curriculum design. The research also has potential application in resource-constrained settings other than in the Sub-Saharan African context.

4.8 Ensuring Rigour

Good or sound research is based on good evidence. There is a need to have standards for the evidence (or quality criteria) to help maintain the rigour required for data collection and the trustworthiness of the data collected, as well as the process of analysis (Krefting, 1991; Miles and Huberman, 1994; Mays and Pope, 1995; Mays and Pope, 2000; Malterud, 2001b; Kuper et al., 2008; Baxter and Jack, 2008; McGloin, 2008; Denscombe, 2010; Frambach et al., 2013; Tavakol and Sandars, 2014b). Strategies are available to ensure rigour is maintained. These include the meticulous recording of data collection and a detailed account of the process of analysis, so that another independent

researcher "could analyse the same data in the same way and come to essentially the same conclusions" (Mays and Pope, 1995:110).

The quality criteria in qualitative research, as summarised by Babbie and Mouton (2001:276) are credibility, transferability, dependability and confirmability. Bearing in mind there may be some overlap between these criteria, they are discussed further in the following sub-sections, with reference to this study.

4.8.1 Credibility

Credibility, or "authenticity" (Miles and Huberman, 1994:278), concerns the extent to which the findings of a study are believable. In Guba's (1981) model, it is termed "truth value". As the data sources in the three case studies were diverse and a range of data collection methods was used, ensuring trustworthiness of the research process (Babbie and Mouton, 2001; Wang et al., 2014) was a challenge. The data analysis that takes place in qualitative research is "a fairly complicated and messy process" (Denscombe, 2010:295) that often requires that the researcher has to deal with numerous anomalies. A variety of procedures was used to try and attain credibility.

Peer debriefing (Babbie and Mouton, 2001) was carried out with a colleague who was outside the context of the study, on an informal basis on average once or twice a month. It was also carried out with my supervisors at our regular meetings every one to two months, as they were often party to reviewing the progress of the study. They provided valuable opportunities to act as a sounding board about ideas and perceptions, and the next steps to take in the study.

The use of triangulation is also part of good research and contributes to ensuring credibility. Data triangulation, the collection of information (data) from different perspectives about the same issue, provided a means for corroboration or validation of the data, and could contribute to the quality and rigour of the data (Mathison, 1988; Thurmond, 2001; Babbie and Mouton, 2001; Denscombe, 2010; Walsh, 2013). As an example, information provided during the interviews with interns at Stellenbosch University and Makerere University on exit-level assessment methods used corroborated the methods stated in documents and actually observed. With methodological triangulation (Mathison, 1988; Thurmond, 2001; Denscombe, 2010), document analysis was used in combination with other methods of data collection such

as observation of assessment events and interviews with examiners (assessors), to allow the findings from one method to be compared with those from another. The interviews with interns allowed for data triangulation of information regarding exit-level assessment methods in their recent experience with the information provided by faculty in their interviews, as well as from document analysis and observations of actual exit-level assessments.

4.8.2 Transferability

Transferability, also termed "fittingness" (Miles and Huberman, 1994:279), refers to "the extent to which the study's findings can be transferred or applied in different settings (Babbie and Mouton, 2001:277). Guba uses the term "applicability" in his proposed model for assessing the trustworthiness of qualitative data (1981). However, this transferability or applicability would appear to be challenged in qualitative research, since "each situation is unique" (Krefting, 1991). Malterud (2001b:485) also cautions that "no study, irrespective of the methods used, can provide findings that are universally transferable". The important point here with regard to transferability is that the onus of judging the transferability of the study findings to other settings does not lie with the researcher but that sufficiently detailed and descriptive data is provided by the researcher in order to allow other researchers to make this judgement (Guba, 1981; Krefting, 1991). To this end, I have attempted to provide as much detail as possible of the research methodology of this study.

As the findings of this study potentially have relevance in other Sub-Saharan African and similar resource-constrained settings, the background information on the research context and as well as the detailed information on the research study design elaborated in earlier sections of this chapter provide transparency of the research process and would contribute to transferability.

4.8.3 Dependability

Dependability is concerned with the consistency of the data, and the extent to which "the findings would be consistent if the enquiry were replicated with the same subjects or in a similar context" (Krefting, 1991:216). Replicating a study would be a way of confirming the consistency of the findings and contributes to rigour. This would require access to detailed descriptions of the research process. In the context of case study

design, Yin (2014:127-128) recommends the setting up of a detailed case study protocol and database, documenting as many steps of the procedures as possible, so that other researchers can follow the procedures. I have tried to provide as much detail as possible with regard to the processes involved in the collection of my data (See also Section 4.4). These include checking transcripts against the original audio recordings for accuracy of transcription, constant comparison of data with the codes to ensure the meaning of the codes has not shifted during the coding process and having an intercoder agreement to cross-check codes (Gibbs, 2007; Creswell, 2014:203).

Transcripts were also sent back to the individual participants and they were invited to check the correctness and interpretation of the data transcribed. Of the 68 interview participants, three indicated at the time of their interviews that it was not necessary for the transcriptions to be sent back to them for checking as they would probably not have the time to read and respond. Twenty interview participants responded to this invitation. The majority agreed with the transcriptions; seven replied with minor changes (such as the clarification of abbreviations and gaps where words could not be heard clearly on the audio recording), and two participants provided detailed clarifications to their grammar and style of language without changing the overall meaning of what they said during their respective interviews.

4.8.4 Confirmability: the Role of Reflexivity

Confirmability refers to "the extent to which the findings are based on the study's participants and settings instead of the researcher's biases" (Babbie and Mouton, 2001:278). Care needs to be taken to ensure that a researcher's previous background and experiences or personal views do not influence the study findings (Tavakol and Sandars, 2014b). However, Denscombe (2010:302-303) argues that in qualitative research, the researcher's own identity or "self" has contributed to the research process and should be regarded as "a crucial resource". With "reflexivity", identifying and declaring the researcher's own stated position openly and frankly does not mean that bias is eliminated, but acknowledges its existence and that it needs to be taken into account (Malterud, 2001b). I therefore provide such an account, as follows:

> *I graduated from medical school some thirty years ago. In my early career, I aspired to be an obstetrician and gynaecologist. But as part of the training, I also spent some time in paediatric rotations, which included general*

paediatrics as well as neonatology. During a six-month rotation on the Neonatal Care Unit of a tertiary teaching hospital, I encountered many babies who were born prematurely or with medical complications as a result of, or following, delivery. Many fortunately survived, with the high-tech care that was available in the unit, but there were those who sadly, could not overcome the overwhelming odds they faced and succumbed. I found these encounters very distressing but learned valuable lessons especially in communication and in dealing with death and dying. I subsequently changed track and have had a very rewarding career in family medicine (general practice), which allowed me to continue my interests in obstetrics and gynaecology and in paediatrics. I was also interested in education, and owe this to my father, who was a general practitioner and a very patient and wonderful teacher. A medical education fellowship in 2005-2006 with the Foundation for the Advancement of International Medical Education and Research (FAIMER)(See Section 3.3), based in Philadelphia PA, USA, was a life-changing experience and provided an opportunity to become part of an international network of health professions educators. My interest in education further led me to pursue postgraduate qualifications in health professions education at masters and now at doctoral level. I also currently hold an academic appointment at a medical school, and in addition to being an examiner at both undergraduate and postgraduate levels, I have been responsible for overall coordination of the Objective Structured Clinical Examinations (OSCEs) at undergraduate exit-level. So this combination of factors and circumstances has all contributed to my wanting to explore an aspect of exit-level assessment in depth and starting this doctoral journey.

With this professional background and training, I considered myself suitably experienced for undertaking the tasks of conducting all the data collection myself, and in particular, the observations of exit-level assessments and the interviewing. However, this professional background as well as my personal experiences may equally have influenced my involvement in the research process. As an example, in my role as the researcher I may have had an influence on faculty with whom I was acquainted prior to the study, in terms of their motivation to participate in the study, or in the responses they provided in

the interviews. I am confident that being mindful of my own perceptions and assumptions, coupled with the measures taken to ensure trustworthiness as described above, have contributed to the rigour of this research endeavour. I remain, however, cognisant of the potential limitations inherent in my own subjectivity and this is acknowledged in Chapter 6 (Section 6.5, Study limitations).

4.9 Concluding remarks

This chapter has defined the research processes used for this study. The purpose and aim of the study have been stated, and the rationale to use case studies has been described and justified. An explanation has also been provided on how the data is to be collected, how it will be analysed and managed. Trustworthiness and rigour of the data collection are important aspects of the validity and reliability of a study and crucial for the research to be credible. The steps that have been put in place to ensure this have been outlined, and my position as a researcher has also been stated. The next chapter concerns what was found at the three study sites pertaining to exit-level assessment.

Chapter 5

Findings

5.1 Introduction

This study represents a multi-facetted approach to investigate the exit-level assessment of clinical competence of medical graduates for the Sub-Saharan African context. In the previous chapter a detailed account of these different facets was provided. The findings presented in this chapter have resulted from the collation and analysis of data collected by document analysis, observations and interviews that took place at the three study sites between April 2014 and June 2015.

Figure 5.1 provides a visual overview of this chapter, relating the questions arising from the study objectives, the methods used to gather data to answer these questions, and the chapter sections under which the data has been collated and analysed.

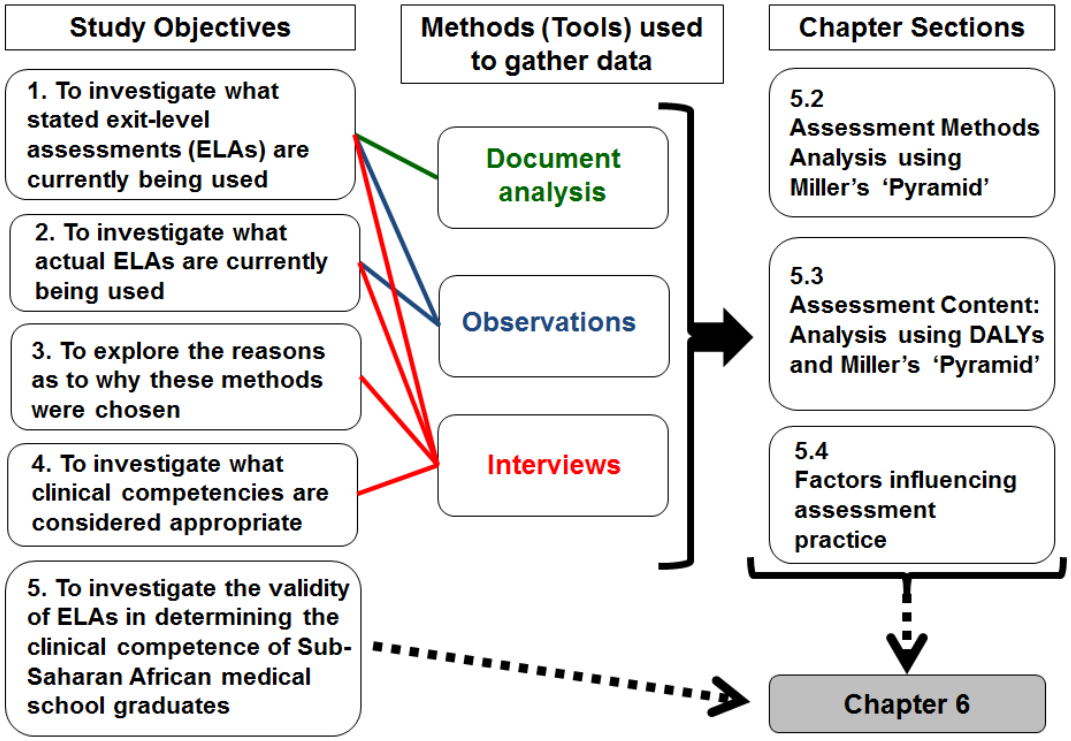


Figure 5.1 Overview of Chapter 5 sections in relation to study objectives

Data gathered from document analysis and clarificatory interviews with clinical clerkship coordinators contributed primarily to the first study objective, which was "To investigate what stated exit-level assessment methods are currently being used in resource-constrained settings". In order to address the second study objective, "To investigate what actual exit-level assessment methods are currently being used in resource-constrained settings", observations were made during the exit-level assessments in Obstetrics & Gynaecology and Paediatrics that took place at the three study sites.

The third and fourth study objectives, "To investigate the reasons for choices of exit-level assessment methods" and "To investigate what clinical competencies are considered appropriate for medical graduates from these medical schools in order to address pertinent health-related issues" respectively, were addressed in interviews conducted with examiners and interns by exploring factors influencing the assessment practices at exit-level. The interviews provided additional information to further elucidate what had been gleaned from the document analysis and the observation of exit-level assessments and also served to triangulate information on what assessment methods were actually being used between the different data sets (Mays and Pope, 2000; Thurmond, 2001; Baxter and Jack, 2008).

Taken as a whole, all data collected helped to build up a picture of assessment methods and content at the three study sites. This will be described in further detail in the following sections, with the discussion and synthesis of all these findings to follow in Chapter 6.

Before moving on to the actual findings of this study and analysis of the data, a brief description is provided of the locations for assessment at the three study sites. The details of venues used at each of the study sites are summarised in Table 5.1.

Table 5.1 Summary of venues where observation of exit-level assessments took place

	University of Botswana	Stellenbosch University	Makerere University
Written assessments	Paeds and O&G: Lecture Hall or Multi-disciplinary laboratory on campus	Paeds: Computer Laboratory on campus O&G: No written assessment	Paeds: Clinical Skills Centre on campus O&G: Physiology Laboratory on campus
Clinical assessments	Paeds and O&G: Multi-purpose teaching rooms on campus	Paeds: Paediatric Outpatient Clinic in Tygerberg hospital O&G: Examination Hall on campus	Paeds: Paediatric Wards at Mulago Hospital O&G: Gynaecology Ward at Mulago Hospital

Abbreviations: O&G = Obstetrics & Gynaecology
Paeds = Paediatrics

The written assessments at the three study sites were generally conducted in venues that could accommodate large numbers of students in one sitting, usually at a regular teaching venue such as a lecture hall, multidisciplinary laboratory, computer laboratory or the clinical skills centre. The clinical assessment for the Department of Obstetrics & Gynaecology at Stellenbosch University was described as an Objective Structured Clinical Examination/Objective Structured Practical Examination (OSCE/OSPE). It did not involve any patients, and was conducted in a large examination hall which was partitioned to enable two OSCE tracks to run concurrently (in parallel). Clinical assessments at University of Botswana, which also took the form of OSCEs, were conducted in multi-purpose teaching rooms in the faculty. Some real patients were used in the Paediatrics and Internal Medicine stations of the medical-based OSCE at the University of Botswana, and arrangements for their attendance as outpatients were made by the respective clinical clerkship coordinators.

The clinical assessments at Stellenbosch University, in both Paediatrics and Obstetrics & Gynaecology were conducted over two days, in order to accommodate the number of students writing the examinations. At the other two study sites, each clinical

assessment for the respective disciplines was conducted in one day. The timeframe of the study was discussed in Section 4.5, and Table 4.6 provided an overview of the timing of data collection at the three study sites.

Clinical assessments that involved real patients were conducted in clinical venues such as hospital wards or in a clinic setting. This has often been the traditional setting for clinical assessments, and in this instance, since the cases used were either inpatients drawn from the wards or outpatients who had been requested to attend specially for the examinations, this would have been a setting that they were familiar with. Standard clinical equipment were made easily available and accessible for use in the assessments as well as for the treatment of patients where required. This was the case for the Department of Paediatrics at Stellenbosch University (which conducted clinical long cases in the Paediatrics outpatient clinic at Tygerberg Hospital), and for both the Departments of Paediatrics and Obstetrics & Gynaecology at Makerere University (which conducted their clinical examinations in their respective hospital wards at Mulago Hospital). Utilising clinical venues which usually provided clinical services meant that the routine services were disrupted temporarily while the examinations took place. This was especially evident with the use of ward space at Mulago Hospital. Two-thirds of a gynaecology ward was used for the Obstetrics & Gynaecology OSCE and clinical long cases. Arrangements were made to accommodate inpatients not involved in the examinations in the remaining ward space or in alternative venues. The paediatric OSCE required three parallel tracks, each track situated on a whole paediatric ward, and this meant the in-patients of those three wards and their carers also required temporary accommodating in alternative venues for the duration of the OSCE.

Real patients were used as cases in the clinical examinations at all three study sites in both Obstetrics & Gynaecology and in Paediatrics, with the exception of the OSCE at Obstetrics & Gynaecology at Stellenbosch University which did not make use of any patients. In addition to real patients, there were also simulated patients in some of the OSCE stations at University of Botswana. Simulated patients are individuals who are trained to play the role of real patients (Collins and Harden, 1998; Cleland et al., 2009).

5.2 Assessment methods

5.2.1 Assessment in the Exit-Level Clinical Clerkships and in the Final Examinations

The exit-level assessment of clinical competence remains a challenge on a number of levels. Careful review of the many factors that influence its practice is needed, centring on the issues of assessment methods and content and also exploring the views of examiners who play a role in assessment. This Section 5.2 provides particulars of the assessment methods used at the three study sites. It will first explore the methods used and offer clarity on the disparate and sometimes confusing terminology used in the three settings. The focus will then shift to the central issue in this section – are the methods employed appropriate to the assessment of clinical competence at exit-level. This is followed by a detailed analysis of the content of the examination questions used in Section 5.3. Section 5.4 then focuses on the interviews that were conducted with examiners, and provides the findings that emerged from qualitative analysis of those interviews.

As a reminder, the observation data described in this study pertains only to the final examinations and not to in clerkship and end of clerkship assessment (This was mentioned earlier in Section 4.4).

At all three sites, assessment during clinical clerkships and during a final examination contribute to students' final marks which means that decisions made about their readiness for practice were based on a range of different assessments. Available documents that were provided for inspection by the clerkship coordinators or respective heads of departments at the study sites included student handbooks or study guides (from Stellenbosch University and the University of Botswana) and information derived from the overall curriculum for the undergraduate medical programme (from Makerere University)(See Section 4.4.1). The analysis of the documents provided data on the exit-level assessment methods used in Obstetrics & Gynaecology and Paediatrics. Overall, the documents across the three study sites show multiple and a diverse range of assessment methods, numbering 21 in total, with 13 of them being used in Obstetrics & Gynaecology and 18 in Paediatrics. The assessment methods are grouped together under three main categories, namely (1) written, (2) performance-based, and (3) other

forms of assessment which did not fall under the previous two categories. This categorisation devised for the preliminary study (Tan et al., 2016) and explained in Section 4.6.2 proved useful in the analysis of the data collected here.

This is summarised in Tables 5.2 and 5.3 for Obstetrics and Gynaecology and Paediatrics respectively together with the percentage contribution of each method to the final mark given that there was variability across the different methods. A range of assessment methods was typically used at exit-level. For the purposes of this study, it was important to understand not only what methods were used and whether those were appropriate to exit-level assessment, but also how heavily the contribution of each was weighed when making decisions at exit-level about students. As an example, it is all good and well that an OSCE is used, but if it only contributes 10% towards a student's final mark, then the way it is used could be called into question. The Stellenbosch University study guides and student handbooks from University of Botswana provided information regarding the assessment methods used, requirements of clinical procedures to be performed or observed and recorded as a logbook, and also included the assessment mark sheets in use. The same level of detailed information was not available from the Makerere University curriculum or clarificatory interviews, as was discussed earlier in Section 4.4.1.

Tables 5.2 and 5.3 show only one instance were a single assessment method was used in the final examination. This was the OSCE/OSPE in Obstetrics & Gynaecology at Stellenbosch University. In all other cases, the final mark was determined by assessment methods from more than one category (This is following the categorisation devised for the preliminary study (Tan et al., 2016) as already explained above. In some cases more than one method were used within a category, varying between one and three (such as the use of multiple choice questions, extended matching questions and short answer questions in written assessment at the University of Botswana in Paediatrics).

Table 5.2 **Assessment methods used during exit-level clinical clerkships and in the final examinations for Obstetrics & Gynaecology at the three study sites, summarised from document analysis and clarificatory interviews**

	Range of assessment methods used during clinical clerkships (% contribution to final mark)			Contribution to final mark	Range of assessment methods used for final examinations (% contribution to final mark)		Contribution to final mark
	Written	Performance-based	Other		Written	Performance-based	
Stellenbosch University	MCQs 15%)	OSCE (10%)	Ward mark (25%)	(50%)	–	OSCE/OSPE (50%)	(50%)
Makerere University	MCQs + SAQs (20%)	Logbook 20%)	Feedback during clinical contact time	(40%)	MCQs (10%) SAQs (10%) Essay (10%)	OSCE (15%) Long case (15%)	(60%)
University of Botswana	–	Mini CEX (5%) Case presentations (10%) Logbook (5%)	PBL Sessions (10%) Seminar presentations (10%)	(40%)	MCQs + EMQs (15%) SAQs (15%)	OSCE (30%)	(60%)

Abbreviations:	PBL = Problem-Based Learning Mini CEX = Mini Clinical Evaluation Exercise MCQs = Multiple Choice Questions EMQs = Extended Matching Questions SAQs = Short Answer Questions OSCE = Objective Structured Clinical Examination OSPE = Objective Structured Practical Examination

Explanatory Notes:
• For further details on individual assessment methods, please refer to the text in Chapters 2 and 5 (Section 5.2.1)

Table 5.3 Assessment methods used during exit-level clinical clerkships and in the final examinations for Paediatrics at the three study sites, summarised from document analysis and clarificatory interviews

	Range of assessment methods used during clinical clerkships (% contribution to final mark)			Contribution to final mark	Range of assessment methods used for final examinations (% contribution to final mark)		Contribution to final mark
	Written	**Performance-based**	**Other**		**Written**	**Performance-based**	
Stellenbosch University	Electronic literature search (5%)	Clinical long case (40%) X-Ray presentation (5%)	Professional conduct & Specialist clinic attendance (satisfactory/ unsatisfactory)	(50%)	"Slide" Written (25%)	Long cases (25%)	(50%)
Makerere University	Case presentations & write-ups (4%) SAQs (20%)	Procedures (4%)	Continuous assessment (12%)	(40%)	MCQs + SAQs (24%)	OSCE (36%)	(60%)
University of Botswana	–	Mini CEX (10%) Case presentations (6%) & Outline management plan (6%) Directly observed procedural skills (4%)	PBL Sessions (10%) Ward evaluation (4%)	(40%)	MCQs + EMQs (15%) SAQs (15%)	OSCE (30%)	(60%)

Abbreviations:	PBL = Problem-Based Learning Mini CEX = Mini Clinical Evaluation Exercise MCQs = Multiple Choice Questions EMQs = Extended Matching Questions SAQs = Short Answer Questions OSCE = Objective Structured Clinical Examination

Explanatory Notes:
- For further details on individual assessment methods, please refer to the text in Chapters 2 and 5 (Section 5.2.1)

In the course of this data analysis, notable differences emerged between some of the assessment methods used in the exit-level clinical clerkships and in the final examinations, and this will be discussed in the following chapter.

5.2.1.1 Written assessment

Written assessment questions were obtained after the scheduled examinations and collated on separate Excel spreadsheets, to summarise information relating to the type of assessment method, number of questions, question topics and the level of Miller's pyramid attained (based on my own clinical judgement and interpretation).

In the **written category**, assessment methods in use included multiple choice questions (MCQs), extended matching questions (EMQs), short answer questions (SAQs) and essays. Explanatory descriptions taken from the literature of these methods have already been provided in Chapter 2 Section 2.4.2. However, this category also included several forms of assessment which have not previously been mentioned, such as electronic literature search, "slide written" test, and case presentations and write-ups, and these will be described in turn. Clarification of some of the terms used (for example, "slide written" test) was obtained at the time of interviews with clinical clerkship coordinators or in subsequent electronic communications with them if the information was not available in the documents analysed.

In an "**electronic literature search**", used by Paediatrics at Stellenbosch University, students are required "to successfully and independently complete an electronic literature search" (as described in the study guide) for relevant literature relating to specific aspects of the clinical condition or disease in one of their clerkship cases. They were then assessed on the presentation of their findings in a "short oral overview" (as described in the study guide) to one of the paediatric consultants. Although the presentation to the consultant was in oral form, students had to write up the findings of their literature search prior to their presentations; and as their write-ups provided more details of their findings than they had time to make in an oral presentation, I placed this in the written assessment category.

The "**slide written**" test, also conducted by Paediatrics at Stellenbosch University, is described in the study guide simply as assessment that "takes the form of written format". On further clarification with the clinical clerkship coordinator, the "slide

written" test follows the format of short answer questions based around clinical scenarios or "patient paper cases" (These are patient cases that are detailed on paper) and/or interpretation of special investigations. It is therefore essentially similar to the short answer question assessment method as described in Chapter 2, in which open-ended questions are structured so that students construct their own concise answers (Webber, 1992; Schuwirth and van der Vleuten, 2003a). Some of the clinical scenarios were accompanied by clinical investigations and/or images which required interpretation as part of the answer. Since images (such as a photographic image or chest radiograph) can be difficult to interpret as part of a paper case, prepared PowerPoint slides illustrating the colour-photographic or digital-radiographic images accompanying these clinical scenarios were used, thereby giving rise to the origin of the term "slide written" test. As later observed during an actual "slide written" test taking place, this examination was conducted in a computer laboratory so that each student had access to an individual desktop computer screen and was able to view the colour-photographic or digital radiograph images accompanying the relevant question, in addition to having a printed copy of the examination paper.

"**Case presentations and write-ups**" used in Paediatrics at Makerere University, as described in the undergraduate medical curriculum, refer to students clerking patients (taking the history and performing a physical examination), then writing up their findings and proposing investigations as well as intervention measures. On further enquiry for clarification, students were marked on both their presentation of their clinical findings (following history taking and physical examination of the patient) as well as their written record of the case which included their plan of management (the proposed investigations and intervention measures). This assessment method therefore straddles both written and performance-based assessment categories, as depicted in Table 5.3.

To sum up, a number of approaches and written assessment methods are used. Multiple choice questions (MCQs) appeared to be used at both the University of Botswana and Makerere University, although less so at Stellenbosch University. Obstetrics and Gynaecology at Stellenbosch did not have a written component in the final examination, and similarly for Paediatrics at Botswana during the clinical clerkships. Paediatrics at Stellenbosch did not make use of MCQs as a written assessment method, but used a

form of short answer questions (the "slide written" examination). The approaches taken, as well as the resources required for these assessment will be discussed later.

5.2.1.2 Performance-based assessment

Performance-based assessment takes place in a clinical context and tests aspects of the complex cognitive and performance skills that are required in the practice of medicine (Swanson et al., 1995; Boursicot et al., 2011). In the final examinations, long cases, objective structured clinical examinations (OSCEs) and objective structured practical examinations (OSPEs) are methods used for assessment. Descriptions of these methods have been provided previously in Chapter 2 Section 2.5. As there was opportunity to observe the conduct of these assessments at the study sites, the details of these observations are also recorded here.

Long cases are used as a performance-based assessment method in Obstetrics & Gynaecology at Makerere University and also in Paediatrics at Stellenbosch University. At both institutions, real patients with clinical signs are used in the long cases, and examiners are present to observe the process and content of the student-patient encounters. At Makerere University, each student is assessed on one long case of 20-minutes' duration by two examiners. Students are observed taking the relevant history from a patient and performing the appropriate physical examination (in the context of Obstetrics & Gynaecology, this usually focuses on an examination of the abdomen) before providing a differential diagnosis to the examiners, followed by discussion of management issues (time permitting). At Stellenbosch University, each student is assessed on two long cases (each of 30-minutes' duration), with one different examiner per case. A similar format is used here in that students are observed performing a physical examination of a child following history taking usually from the child's care-giver (if present), discussion of the differential diagnosis and management issues in the case.

A second performance-based assessment method, used at all three study sites, is the **objective structured clinical examination** (OSCE). At University of Botswana, where the approach to assessment is described as 'integrated' (as mentioned in the Paediatric student handbook), the "medical-based" OSCE comprises questions from Paediatrics as well as other medical-based disciplines (Internal Medicine and Family Medicine), and the Obstetrics & Gynaecology contribute questions to the "surgical-based" OSCE (which

also includes questions from General Surgery, Vascular Surgery, Orthopaedic Surgery, Ophthalmology, Otorhinolaryngology, Anaesthesiology and Emergency Medicine). The stations are set and marked by examiners from the respective disciplines. Paediatrics and Obstetrics & Gynaecology at Makerere University conduct their respective discipline-based OSCEs.

At Stellenbosch University the Obstetrics & Gynaecology "**OSCE/OSPE**" comprises a total of 16 stations, each 7-minutes' long. The eight OSCE stations focus on clinical scenarios. Examiners are present at three of these stations and require students to perform an abdominal or pelvic examination on a mannequin (model), or some aspect of communications skills (such as taking consent prior to physical examination and/or explanation of a clinical examination procedure) with the station examiner role playing a patient. Examiners are not present at the remaining five OSCE stations where students provide written answers to questions such as data interpretation or the management of clinical problems, again essentially a short answer question assessment format as previously described (Refer to Appendix Q for details). Each of the four stations in the OSPE part of the examination comprises 7-minutes preparation time for the student in a station followed by 7-minutes assessment time with an examiner at the following station (Therefore although each OSPE question in effect has two stations, for the purposes of further analysis and discussion in this study, I counted these as four stations). The OSPE station questions focus on the clinical management of obstetric or gynaecological conditions or situations. The student-examiner interactions at these stations are of an oral nature. From these observations, it appears that these OSPE stations in the Obstetrics & Gynaecology OSCE/OSPE are essentially structure orals, perhaps more in way of "cognitive skills" rather than procedural or clinical skills stations. The duration of OSCE stations at Makerere University and University of Botswana is 10-minutes; at Stellenbosch University, the stations are 7-minutes' long.

Other methods of performance-based assessment in use during the clinical clerkships include the "mini CEX" (or Mini Clinical Evaluation Exercise), "case presentations", "outline management plan", "logbooks", "directly observed procedures" and "procedures", and "X-Ray presentations" (Refer back to Tables 5.2 and 5.3). Both Obstetrics & Gynaecology and Paediatrics at University of Botswana make use of the mini CEX. This assessment method has already been described in Chapter 2 Section 2.5.

Brief descriptions of the other methods mentioned here are given below, based on available information in the documents analysed. The choice of these other methods of performance-based assessment varies among the study sites and there does not appear to be any one method used more often than another. These other methods appeared to be "minor" methods, since their contribution to the final mark was mostly in the 4 to 10% range, compared to the larger percentages with OSCE and long cases.

Both Obstetrics & Gynaecology and Paediatrics at University of Botswana also make use of "**case presentations**" and, in addition, students in Paediatrics are required to provide an "**outline management plan**". Student handbooks provide detailed information and the format of the assessment mark sheet to be used for clinical presentations. These presentations are done in a clinical setting, such as on a ward round. Students are marked on their clinical findings obtained on clerking a patient (history taking and physical examination of the patient), their clinical judgement (encompassing a differential diagnosis of the case, appropriate investigations and a management plan), their interaction with the patient and presentation skills (in presenting their findings to faculty or a member of the medical team) and aspects of professionalism (for example, this is described in the handbooks as "[Being] respectful and establishes trust", and "Attends to patient's comfort and respects modesty"). In Obstetrics & Gynaecology, marks are given for each of these components and combined to give one overall score, whereas in Paediatrics, marks for clerking and presentation are awarded as one score, and marks for the clinical judgement component are awarded as a separate score under the term "outline management plan". For reference, the University of Botswana Paediatrics case-presentation mark sheet is reproduced in Appendix R. The format of the case presentation in use at University of Botswana, as described here, appears to be similar to that used in Paediatrics at Makerere University, as described in Section 5.2.1.1. However, there is no reference in the University of Botswana student handbook to students being marked on their written record of the case as they do at Makerere University, so this assessment method was therefore allocated to the performance-based category.

The University of Botswana Obstetrics & Gynaecology "**logbook**" refers to clinical procedures that students are required to assist in or perform under supervision, together with a specified number of procedures as the minimum requirement. Each

procedure that the student performs is recorded in the logbook and countersigned by a clinician supervising the procedure. For example, students are required to perform 10 normal vaginal deliveries, perform and repair 10 episiotomies, assist in 2 Caesarean section deliveries, assist in 5 major operations (such as a total abdominal hysterectomy or a laparotomy) during their rotation. A "logbook" is also used in Obstetrics & Gynaecology at Makerere University "to assess skills acquisition" (as stated in the MBChB degree curriculum). However, there was insufficient information from the Makerere University documents to provide a further detailed breakdown of the skills assessed.

Students at two of the institutions were formally required to perform procedural skills, but the approach to assessing these varied considerably. At the more formal end of the spectrum, the University of Botswana required students to perform and successfully complete certain procedural skills before they could be signed off by a supervising clinician. The University of Botswana Paediatrics student handbook provides details of these "**Directly Observed Procedural Skills**" (**DOPS**). They include non-invasive skills (such as Apgar scoring of a new-born infant, interpretation of chest radiograph and electrocardiogram results) and invasive skills (such as intravenous cannulation and phlebotomy, bladder catheterisation) performed on patients, procedures performed on a mannequin or model (for example, endotracheal-intubation and intraosseous-line placement), and procedures that students need to observe (such as neonatal or paediatric resuscitation, lumbar puncture, and obtaining informed consent). Detailed, structured assessment forms are also provided in the student handbook for the assessment of DOPS, and an example of one such DOPS assessment form is displayed as Appendix S.

At the less formal end of the scale, students in Paediatrics at Makerere University also conduct what is described in the student handbook "**procedures**" (listed in the curriculum as venepuncture, lumbar puncture, the insertion of intravenous infusion lines and nasogastric tubes, immunisations). These procedures are performed on ward patients to obtain clinical samples for laboratory-based tests (such as urine and stool examination, and blood-film preparation and examination). There was no clear explanation provided in the relevant student information documents on how these

procedures are assessed, and no added information was obtained on this issue from clinical clerkship coordinators.

For the "**XRay presentation**" conducted in Paediatrics at Stellenbosch University, students are required to present a chest radiograph ("XRay") of a neonate or child to the responsible consultant and are assessed according to the structured assessment mark sheet in the study guide. In the process, students learn a systematic approach to presenting the findings of this clinical investigation, and also a process of data interpretation. It could be argued that an X--Ray presentation such as this could be adapted into an OSCE station, and with a live or standardised patient present, so that the student not only had to interpret data (a knowledge application process) but could also be assessed on the demonstration of a skill, such as patient explanation. This would emphasize the application of knowledge and skills learnt in a classroom setting to a more clinical one.

In summary, as with the written assessment methods, a diverse range has been seen in the performance-based assessment category. The approaches, feasibility, resources required, all need to be considered with reference to the context, and this will be discussed further later.

5.2.1.3 Other assessment

A number of assessment methods in the "Other" assessment category, featured notably in exit-level clinical clerkship assessment as described in Section 5.2.1 but not in the final examinations (Refer back to Tables 5.2 and 5.3). These included:

"**ward mark**" in Obstetrics & Gynaecology at Stellenbosch University,

"**feedback during clinical contact time**" in Obstetrics & Gynaecology at Makerere University,

"**ward evaluation**" in Paediatrics at University of Botswana,

"**professional conduct**" and "**specialist clinic attendance**" in Paediatrics at Stellenbosch University, and "**continuous assessment**" in Paediatrics at Makerere University.

These methods were described in the respective institutional documents as dealing with various aspects of professionalism, attitude and conduct. The "ward mark" in

Obstetrics & Gynaecology at Stellenbosch University (as described in the study guide) related to marks that were allocated based on ward work, student interaction with patients, ward staff and peers, preparation for ward rounds and general impression. The precise manner of and criteria for assessment were unclear and there was little or no further details available in the documents available for analysis. The "ward professionalism evaluation form" was available for reference in the Botswana Paediatrics student handbook and provided a breakdown of aspects of professional behaviour that were assessed and contributed to ward evaluation. No added information was obtained on this matter during the interviews with examiners as this was not explored further at the time of the interviews.

A problem-based learning (PBL) approach was used as a method of teaching at University of Botswana, and "**PBL Sessions**" are conducted both in Obstetrics & Gynaecology and in Paediatrics. A fixed list of topics is provided in the respective student handbooks (examples include HIV and prevention of mother to child transmission, medical disorders of pregnancy, postmenopausal bleeding, and evaluation of ovarian tumours), and students are assessed on their "participation and knowledge" during the PBL sessions using structured detailed mark sheets provided for reference in the student handbooks. PBL is also used as a method of teaching at Makerere University. In the Paediatrics curriculum document a list of topics is provided as well as marking criteria (based on "time management, participation, social congruence, interpersonal skills, communication skills and self-assessment"). Although there is no clear indication in the curriculum documents and further details were not obtained during interviews with examiners, marks from the PBL sessions at Makerere University most likely form part of "continuous assessment" and this is therefore allocated to the "Other" assessment category.

In Obstetrics and Gynaecology at University of Botswana, students prepare and conduct weekly "**seminar presentations**" on selected topics (examples include induction and augmentation of labour, multiple pregnancy, family planning and contraceptive methods, cervical cancer screening, evaluation and management of sexual assault). Students are assessed according to the structured mark sheet which is available for reference in the student handbook. The marking criteria are based on the disease or disorder topic, clinical management issues, primary literature review, organisation and

flow of the discussion around the topic, knowledge, and aspects of communication skills including verbal and non-verbal communication, timing of the presentation and the PowerPoint slide presentation. However, such topic presentations did not form part of the data collected for this study, and it is possible or even likely, that there was assessment of relevant topics which were not accessible to me as a researcher. This would have implications for any studies of the "whole curriculum", as ensuring the gathering of all relevant data would be an enormous challenge.

Overall, a range of assessment methods were used at exit-level, and the observed exit-level assessment methods matched the methods stated in institutional documentation. The amount of detail captured in Tables 5.2 and 5.3 therefore emphasised both the diversity as well as the variance in terms of combinations across institutions. This variance is also reflected in the way in which marks were allocated and this will be described in the next Section 5.2.2.

5.2.2 Allocation of marks

Thus far we have looked at the range of methods used, but claiming range or diversity is of no value if, for example, one method dominates in terms of the extent to which it contributes to the final mark and therefore the 'judgement' being made about the student's competence. For this reason, this next section explores the weighting that was allocated to each different assessment type.

Marks from the assessment methods used during exit-level clinical clerkships as well as the final examinations both contributed to the final mark, and the proportion for each of these two components is shown (as a percentage) in Tables 5.2 and 5.3 for Obstetrics & Gynaecology and Paediatrics respectively. A breakdown of the allocated marks is also indicated in the tables. Within each study site, the two disciplines followed the same percentages toward the final mark: at Stellenbosch University, assessment during the clinical clerkships and in the final examinations contributed 50% equally. At Makerere University and University of Botswana, the contributions to the final mark were correspondingly 40% and 60% for the clinical clerkship assessment and the final examinations.

The marks allocated for individual assessment methods used during clinical clerkships varied between 4 and 40%, compared to 10 and 50% for methods used in the final

examinations. For example, logbooks and mini CEX used in Obstetrics & Gynaecology clinical clerkship assessment at the University of Botswana contributed 5% each, whereas MCQs and EMQs in the final examination were allocated 15% and OSCE 30%. In another instance, mini CEX used in Paediatrics clinical clerkship assessment at the University of Botswana was allocated 10% of the final mark, with 4% for DOPS and 6% each for case presentations and outline management plans. These differences in allocation of marks could also serve as a signal to students on the degree of importance attached to respective assessment components. Students might selectively focus their attention to assessment methods that carry a higher weighting in an attempt to maximise their final marks. There was no information in the institutional documentation regarding the rationale for the percentages of marks contributing toward the final mark, and this issue was not explored further at the time of interviews with examiners. However, this issue of weighting of marks will be discussed again in Chapter 6.

The percentage of marks allocated to the "written", "performance-based" and "other" categories was also variable. At University of Botswana, clinical clerkship marks were determined by "performance-based" and "other" categories, and given equal weighting in Obstetrics & Gynaecology; but in Paediatrics, the proportion of marks was 65:35. At Stellenbosch University, weightings of the three categories of assessment methods for Obstetrics & Gynaecology resulted in the proportion of marks for "written" to "performance-based" to "other" being split 30:20:50. Paediatrics at Stellenbosch University assessed students on professional conduct and specialist clinic attendance in the "other" category as either satisfactory or unsatisfactory instead of a numeric mark. The "written" marks in proportion to the "performance-based" marks were 10:90. Makerere University Obstetrics & Gynaecology allocated equal percentages of marks to the "written" and "performance-based" categories of assessment; there was also assessment in the "other" category in the form of "feedback during clinical contact time" but however, there was insufficient information in the curriculum document to indicate if numeric marks were awarded for this. Paediatrics at Makerere University also utilised assessment methods in all three categories with the proportion of marks being split 55:15:30. Immediately evident, therefore, is the fact that some students are being deemed competent based predominantly on their 'performance' (as seen at

Stellenbosch University), while others predominantly on their ability to demonstrate their knowledge (as seen Stellenbosch University and Makerere University).

Turning to the final examinations, the percentage of marks allocated for "written" and "performance-based" categories of assessment methods carried equal weighting at all three study sites, with the exception of Paediatrics at Makerere University, where the written marks in proportion to the performance-based marks were 40:60.

Table 5.4 Summary of the number of questions and percentage contribution to the final mark for each assessment method used in the final examinations

	Obstetrics & Gynaecology		**Paediatrics**	
	Total number of questions	*Percentage of final mark*	*Total number of questions*	*Percentage of final mark*
Stellenbosch University	OSCE/OSPE (12)	(50%)	Slide written (19) Long Case (2)	(25%) (25%)
Makerere University	MCQ (14) SAQ (10) Essay (2) OSCE (6) Long Case (1)	(10%) (10%) (10%) (15%) (15%)	MCQ (50) SAQ (4) OSCE (12)	} (24%) } (36%)
University of Botswana	MCQ (12) EMQ (5) SAQ (5) OSCE (4)	} (15%) } (15%) (30%)	MCQ (12) EMQ (6) SAQ (5) OSCE (5)	} (15%) } (15%) (30%)

Abbreviations:
MCQ = Multiple Choice Questions
EMQ = Extended Matching Questions
SAQ = Short Answer Questions
OSCE = Objective Structured Clinical Examination
OSPE = Objective Structured Practical Examination

As can be seen from the table, while OSCE is used at all three institutions, the contribution is limited in most instances. Only the Stellenbosch University Obstetrics & Gynaecology OSCE/OSPE and the Makerere University Paediatrics OSCE exams have enough stations to yield data that might be reliable although there is no agreed consensus on the minimum required number of OSCE stations. Even though MCQs might be used well (see Section 5.3.1 for a comment on the appropriateness of how MCQs are used to exit-level assessment) in only one instance are sufficient MCQs (in Makerere University Paediatrics) used to make any claim of content validity possible (although similarly, there is no agreed consensus on the number of MCQs required to

deem the assessment valid and reliable). These issues will be addressed again in Chapter 6 as the findings of this study are synthesized.

To recap, these two Sub-Sections 5.2.1 and 5.2.2 have provided detailed descriptions of the data derived from document analysis supplemented by clarificatory interviews, and have clearly demonstrated the variance in the allocation of marks for the range of assessment methods at different levels: across different categories of assessment methods, between the clinical disciplines under study, and also across the study sites. The upcoming sections deal with analysis of the assessment methods.

5.2.3 Analysis of assessment methods according to Miller's 'Pyramid'

The focus of this study is to determine the validity of exit-level assessment in relation to the clinical competence required of medical graduates for selected health issues in the Sub-Saharan Africa context. The key to this investigation, therefore, is the extent to which relevant evidence regarding the students' clinical competence can be generated. In order to do this, however, there is a need to look at how exit-level assessment in medical schools takes place and how judgements are made about clinical competence. As described in Chapter 2, competence ought to be demonstrated rather than described. To this end, I adopted Miller's 'pyramid' (Refer to Section 2.4.1) as an analytical framework to explore the appropriateness of the methods used for the purpose of making decisions about clinical competence. To recap, Miller's 'pyramid' is illustrated in Figure 5.2, using Miller's original descriptions of each tier of the 'pyramid', what is measured at each tier, and examples of assessment methods.

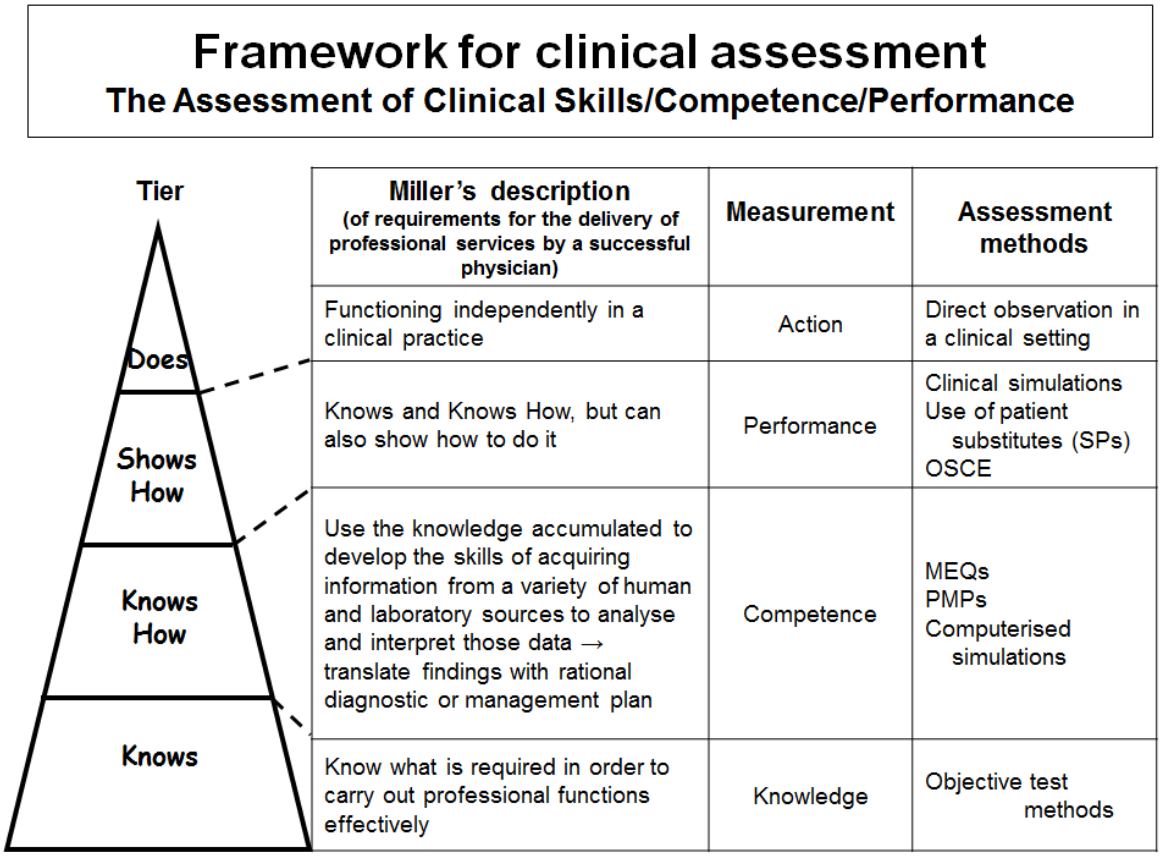


Figure 5.2 Framework for Clinical Assessment according to Miller (1990)

Figure adapted from Miller (1990)

(Refer back to Figure 2.1 for explanation and notes on abbreviations)

If the assessment methods used in the final examinations at the three study sites were to be classified according to Miller's original descriptions, based on what the assessment method typically requires of the student, the arrangement would take on the appearance in Table 5.5. However, this would only work on the assumption that the same tier of Miller's 'pyramid' applies to all individual questions in each assessment method. Actual evidence if this is the case needs to be sought by analysing the content of each individual question in relation to the tiers. It is therefore essential to understand the process of this content analysis first, before returning to the subject of analysis of assessment methods according to Miller's 'pyramid' (which will be addressed further in Section 5.3.1).

Table 5.5 Summary of assessment methods in the final examinations

	Obstetrics & Gynaecology				Paediatrics			
	K	**KH**	**SH**	**D**	**K**	**KH**	**SH**	**D**
Stellenbosch University	-	-	OSCE/OSPE	-		Slide written	Long case	-
Makerere University	MCQ Essay	SAQ	OSCE Long case	-	MCQ	SAQ	OSCE	-
University of Botswana	MCQ	EMQ SAQ	OSCE	-	MCQ	EMQ SAQ	OSCE	-

Abbreviations:	MCQ = Multiple Choice Questions EMQ = Extended Matching Questions SAQ = Short Answer Questions OSCE = Objective Structured Clinical Examination OSPE = Objective Structured Practical Examination	K = Knows KH = Knows How SH = Shows How D = Does

5.3 Assessment content

To take a stance regarding the validity of assessment for regional needs, two questions needed answering: (1) Was the content assessed relevant to regional health needs?, and (2) Was that being assessed – knowledge or competence – relevant to exit-level assessment of graduates in the region? A specific focus within this study in determining the extent to which exit-level assessment at selected Sub-Saharan African medical schools can provide evidence of clinical competence of medical graduates had to do with the appropriateness of such assessments to respond to health needs on the continent. Therefore, in addition to investigating the exit-level assessment methods used in the three study sites, the content of the exit-level assessment also required some attention as this was a source of evidence towards establishing validity. Observation of exit-level assessments at the three study sites provided information on the actual process of assessment that took place. Details of the question content were obtained from the written examination papers as well as during the conduct of the performance-based

assessment. The process of conducting the observations has already been described in detail in Chapter 4 (See Section 4.4.2).

There were several steps in the analysis of assessment content, and these will be described in the remainder of this section. As a first step, the written examination papers were analysed in terms of format and form. This process was also repeated with the performance-based assessments that took place. I conducted this analysis separately per study site and per discipline. The Stellenbosch University Paediatrics "slide" written paper is used here as an example to illustrate the steps of analysis.

As could be seen from Table 5.3, Stellenbosch University only made use of one form of written assessment as part of the final component for Paediatrics, namely the "slide written" examination paper. During analysis, it was noted that this paper actually had 13 questions. Twelve questions each had a clinical scenario followed by several sub-questions focussing on aspects of a particular clinical topic and requiring short answers. The 13th question had seven sub-questions also in a short answer format and some with accompanying photograph or radiograph image, but each sub-question had its own clinical scenario and a separate clinical topic. In essence, each sub-question could be considered a stand-alone question. For the purposes of further analysis and discussion in this study, therefore, the total number of questions in the "slide written" examination paper was taken as 19.

As mentioned in Section 4.6.2, the topics or themes of individual questions in each assessment method were tabulated as a first part of the analytical process. This facilitated the generation of a detailed account of the content included in the many different assessment events that students were exposed to. Table 5.6 illustrates the question topics in the Stellenbosch University Paediatrics written final examination paper:

Table 5.6 Question topics in the Paediatrics written final examination paper at Stellenbosch University

Question No.	Topic
1	Acute lower respiratory tract infection
2	Trisomy 21 (Down's syndrome)
3	Neurofibromatosis
4	Interpretation of Electrocardiograms & Resuscitation
5	Thrombocytopenia/Idiopathic Thrombocytopenic Purpura (ITP)
6	Nephrotic syndrome
7	Congenital adrenal hyperplasia
8	Asthma; atopy
9	HIV positive mother; Infant feeding practices
10	Management: Premature baby with apnoea; HIV positive mother
11	Duchenne's muscular dystrophy; motor development
12	Tuberculosis meningitis
13	Cystic fibrosis
14	Diabetic lipohypertrophy
15	Protein energy malnutrition
16	Diarrhoea & Vomiting: Interpretation of blood gases and electrolytes
17	Biliary atresia
18	Tetralogy of Fallot
19	Rash: Scarlet fever

This data was then used in the second phase of analysis, and the findings from this analysis are described next.

The analysis of the content of the exit-level assessments was done using two analytical tools, the first being the Framework for Clinical Assessment according to Miller (1990) (or Miller's 'pyramid') as mentioned above, and the second, the causes of Disability Adjusted Life Years (DALYs) for disease conditions for the Sub-Saharan African regions relevant to this study (As mentioned earlier in Section 4.6.2, DALYs replaced Millennium Development Goals (MDGs) because the latter were more suitable at the level of overall global health policies but not specific enough to inform at the level of curriculum and assessment).

5.3.1 Analysis of appropriateness of content using Miller's 'pyramid'

The first tool used in the second phase of the analysis of the content of the exit-level assessments looked into the appropriateness of the content for exit-level. Drawing from the data collected by document analysis and observations, and with reference to the

descriptions that Miller (1990) used for each tier, every assessment question was assigned to an appropriate tier of the 'pyramid'. Deductive coding was applied in this process, as discussed in Section 4.6.2, and several examples are given here to demonstrate this.

In the first example, the question topic in question 12 (provided in Box 1 below) related to tuberculosis meningitis as the most likely diagnosis (This was also indicated in the model answer that was provided). In order to arrive at this diagnosis, the student would have to draw upon prior knowledge and interpret all the information provided in the clinical vignette as well as the physical examination and investigation findings. I therefore assigned this question to the "Knows How" tier of Miller's 'pyramid'. As mentioned in Section 4.6.2, the process of allocation of question topics to the appropriate tier of Miller's 'pyramid' was reviewed by one study supervisor who was a qualified medical practitioner with considerable clinical experience of the Sub-Saharan African context.

Box 1

["Knows How"]

Stellenbosch University, Paediatrics
Question 12

An 18-month-old boy presents to you at primary-level hospital with a one-week history of not wanting to play, being sleepy and vomiting once or twice a day for the past 3 days. The child was seen by the clinic 3 days previously with fever and was given amoxicillin, which made no difference.
On the Road-to-Health booklet the child has loss of weight for the past 3 months.

On examination the child seems listless and irritable, but is awake. He has a temperature of 38ºC with normal ear-nose-throat examination. While being examined, the child has a generalised convulsion which lasts 5 minutes.
Other systems examine normally and the urine examination (dipstix) is clear. The blood glucose done in the outpatients departments registers in the normal range.
The child wakes up after the seizure, with no focal neurological signs on examination.

A lumbar puncture is done with CSF results as follows:

Clear
30 polymorphonuclear leucocytes/mm3
220 lymphocytes/mm3
Total protein 1/2 g/dL
Glucose 1.8 mmol/L (Serum glucose 5 mmol/L).

(a) What is the MOST LIKELY diagnosis?
(b) What medication would you prescribe for this child?
(c) What additional information and/or special investigations would you do to confirm your clinical suspicion? Name FIVE.

["Knows"]

Stellenbosch University, Paediatrics
Question 13

A 6 month-old infant with respiratory problems is a member of a family with a brother who has been diagnosed with cystic fibrosis (CF).

(a) Name TWO tests that you would do to exclude CF in this baby.
(b) What genetic test would you do to confirm the diagnosis and which are the common mutations in South Africa?

The second example in Box 1, question 13, related to cystic fibrosis. The question tests factual information regarding specific investigations that would help confirm a diagnosis of cystic fibrosis. My interpretation here was that this is a question testing knowledge and knowledge recall, and I therefore assigned this question to the "Knows" tier of Miller's 'pyramid'.

As a next step, each examination question was assigned to a tier of Miller's 'pyramid' and tabulated. The resultant table showing each question topic in the Stellenbosch University Paediatrics written final examination paper and the corresponding allocated Miller's 'pyramid' tier is illustrated in Table 5.7.

Table 5.7 Question topics in the Paediatrics written final examination paper at Stellenbosch University and the corresponding allocated tier of Miller's 'pyramid' using deductive coding

Question No.	Topic	Miller's 'pyramid' tier			
		K	KH	SH	D
1	Acute lower respiratory tract infection				
2	Trisomy 21 (Down's syndrome)				
3	Neurofibromatosis				
4	Interpretation of Electrocardiograms & Resuscitation				
5	Thrombocytopenia/Idiopathic Thrombocytopenic Purpura (ITP)				
6	Nephrotic syndrome				
7	Congenital adrenal hyperplasia				
8	Asthma; atopy				
9	HIV positive mother; Infant feeding practices				
10	Management: Premature baby with apnoea; HIV positive mother				
11	Duchenne's muscular dystrophy; motor development				
12	Tuberculosis meningitis				
13	Cystic fibrosis				
14	Diabetic lipohypertrophy				
15	Protein energy malnutrition				
16	Diarrhoea & Vomiting: Interpretation of blood gases and electrolytes				
17	Biliary atresia				
18	Tetralogy of Fallot				
19	Rash: Scarlet fever				

Abbreviations: K = Knows; KH = Knows How; SH = Shows How; D = Does

This process of analysis was applied in turn to all the questions or question topics from the written assessments as well as the performance-based assessments used (long cases and OSCE) in the final examinations at the three study sites (thus 186 analyses in total). Examples of actual questions in full are provided from each written assessment method as well as performance-based assessments used (long cases and OSCE), and their allocation to the respective tier of Miller's 'pyramid', are provided in Appendix T. For further reference the full detailed breakdown of all the topics or themes of individual questions in each assessment method is provided in Appendices Q (Stellenbosch University), U (University of Botswana) and V (Makerere University).

5.3.1.1 Synthesis of method, content and level

At this point, taking cognisance of the description of the process of content analysis that was provided in Section 4.6.2, there is a need to return briefly to the previous Sub-

Section 5.2.3 to complete the findings obtained from analysis of assessment methods according to Miller's 'pyramid'. With each question now assigned to a tier of the 'pyramid', a more detailed picture emerges of the distribution of questions across the tiers, showing shifts from the initial mapping as noted in Table 5.5. This is summarised in Tables 5.8 and 5.9, with the number of questions (and percentage contribution to the final mark) in relation to the tiers of Miller's 'pyramid' for each assessment method used in the final examinations at the three study sites for Obstetrics & Gynaecology and Paediatrics respectively (As noted in Section 4.4, detailed access to assessment content was only possible for the final examinations and not for in-course assessment. As a result, in the following Tables 5.8 and 5.9, only the analysis for the final examinations is presented).

Overall, the questions from the **written assessment** methods lay in the "Knows" and "Knows How" tiers of Miller's 'pyramid' for both Obstetrics & Gynaecology and Paediatrics. "Slide written" test questions from Paediatrics at Stellenbosch University have already been discussed and examples provided to illustrate differences and how the questions were allocated to the "Knows" (10.5% of the final mark) and "Knows How" (14.5%) tiers.

Table 5.8 **Summary of the number of questions and percentage contribution to the final mark for each assessment method used in the final examinations for Obstetrics & Gynaecology at the three study sites in relation to the tiers of Miller's 'pyramid'.**

			Tiers of Miller's 'pyramid'						
			KNOWS		**KNOWS HOW**		**SHOWS HOW**		**DOES**
	Assessment method and total number of questions	*Percentage of final mark*		*Percentage of final mark*		*Percentage of final mark*		*Percentage of final mark*	
Stellenbosch University	OSCE/OSPE (12)	(50%)	OSCE/OSPE (1)	(4.2%)	OSCE/OSPE (8)	(33.3%)	OSCE/OSPE (3)	(12.5%)	-
			Total	4.2%	*Total*	33.3%	*Total*	12.5%	
Makerere University	MCQ (14) SAQ (10) Essay (2) OSCE (6) Long Case (1)	(10%) (10%) (10%) (15%) (15%)	MCQ (14) SAQs (2)	(10.0%) (2.0%)	SAQs (8) Essay (2) OSCE (2)	(8.0%) (10.0%) (5.0%)	OSCE (4) Long Case (1)	(10.0%) (15.0%)	-
			Total	12.0%	*Total*	23.0%	*Total*	25.0%	
University of Botswana	MCQ (12) EMQ (5) SAQ (5) OSCE (4)	} (15%) } (15%) (30%)	MCQ (6)	(5.3%)	MCQ (6) EMQ (5) SAQ (5)	(5.3%) (4.4%) (15%)	OSCE (4)	(30%)	
			Total	5.3%	*Total*	24.7%	*Total*	30%	

Abbreviations:	MCQ = Multiple Choice Questions EMQ = Extended Matching Questions SAQ = Short Answer Questions OSCE = Objective Structured Clinical Examination OSPE = Objective Structured Practical Examination	**Explanatory Notes:** • Figures in parentheses refer to the number of questions and percentage contribution to the final mark

Table 5.9 Summary of the number of questions and percentage contribution to the final mark for each assessment method used in the final examinations for Paediatrics at the three study sites in relation to the tiers of Miller's 'pyramid'.

		Tiers of Miller's 'pyramid'			
		KNOWS	**KNOWS HOW**	**SHOWS HOW**	**DOES**
	Assessment method and total number of questions / *Percentage of final mark*	*Percentage of final mark*	*Percentage of final mark*	*Percentage of final mark*	
Stellenbosch University	Slide written (19) (25%) Long Case (2) (25%)	Slide written (8) (10.5%)	Slide written (11) (14.5%)	Long Case (2) (25.0%)	-
		Total 10.5%	*Total* 14.5%	*Total* 25.0%	
Makerere University	MCQ (50) } (24%) SAQ (4) } OSCE (12) (36%)	MCQ (50) }(22.7%) SAQ (1) } OSCE (3) (9.0%)	SAQ (3) (1.3%) OSCE (2) (6.0%)	OSCE (7) (21.0%)	-
		Total 31.7%	*Total* 7.3%	*Total* 21.0%	
University of Botswana	MCQ (12) } (15%) EMQ (6) } SAQ (5) (15%) OSCE (5) (30%)	MCQ (1) (0.8%)	MCQ (11) }(14.2%) EMQ (6) } SAQ (5) (15.0%)	OSCE (5) (30.0%)	
		Total 0.8%	*Total* 29.2%	*Total* 30.0%	

Abbreviations:	
	MCQ = Multiple Choice Questions EMQ = Extended Matching Questions SAQ = Short Answer Questions OSCE = Objective Structured Clinical Examination OSPE = Objective Structured Practical Examination

Explanatory Notes:
- Figures in parentheses refer to the number of questions and percentage contribution to the final mark

Multiple choice questions (MCQs) were used at both Makerere University and the University of Botswana in both disciplines. Each question was allocated to the corresponding Miller's 'pyramid' tier following the process described previously with the "slide written" questions from Paediatrics in Stellenbosch University. The MCQs at Makerere University in both disciplines tested factual knowledge and knowledge recall, and all remained in the "Knows" tier. Some question topics are shown, as examples, in Box 2 below.

Box 2

["Knows"]

Makerere University, Obstetrics & Gynaecology

(Question topics)
- Sickle cell disease in pregnancy
- Therapy in HIV positive mothers

["Knows"]

Makerere University, Paediatrics

(Question topics)
- Clinical manifestations of malaria
- Normal developmental milestones at a routine child health clinic check

Even though the use of a much larger number of MCQs than elsewhere suggests better content coverage or content validity, all questions were at the "Knows" tier. This is an illustration of how content might be appropriate to regional needs but the level of assessment not appropriate to exit-level assessment.

At University of Botswana, the Obstetrics & Gynaecology MCQs were equally divided between the "Knows" and "Knows How" tiers. Question 14, as shown in Box 3, tests factual knowledge only with regard to staging in a molar pregnancy ("Knows" tier), whereas in Question 17, the appropriate choice of contraceptive method requires the student to refer to the information provided in the clinical vignette and apply previous knowledge regarding contraceptive methods. This therefore supports the allocation of this question to the "Knows How" tier.

Box 3

["Knows"]

University of Botswana, Obstetrics & Gynaecology
Question 14

A 41-year-old, Gravida 3 Para 2 Abortion 1, was managed for molar pregnancy 2 months ago. Now she presents with excessive per vaginal bleeding and has a secondary lesion on the vaginal wall.
Serum βhCG titre is 70,000 IU/ml (Normal <5 βhCG mIU/ml).
Computerised tomography reports metastases to the lungs.

Which one of the following best describes her disorder?

A. Non-metastatic gestational trophoblastic neoplasia
B. Metastatic gestational trophoblastic neoplasia stage I
C. Metastatic gestational trophoblastic neoplasia stage II
D. Metastatic gestational trophoblastic neoplasia stage III
E. Metastatic gestational trophoblastic neoplasia stage IV

["Knows How"]

University of Botswana, Obstetrics & Gynaecology
Question 17

A 32-year-old, Gravida 1 Para 1 and HIV negative, comes for her postnatal visit. She has chronic hypertension with a Body Mass Index of 30kg/m^2 and BP of 149/95 mmHg. She is not breastfeeding.

Which of the following contraceptive methods is most suitable for her?

A. Dermal patch
B. Vaginal ring
C. Cu T 380A
D. Bilateral tubal ligation
E. Combined oral contraceptive pill

The majority of the University of Botswana Paediatric MCQs were in the "Knows How" tier. However, Questions 62 and 69 are shown here in Box 4 to illustrate their allocation to the "Knows How" and "Knows" tiers respectively. Question 62 requires interpretation of the history and physical findings provided in order to reach the most likely diagnosis.

Box 4

["Knows How"]

University of Botswana, Paediatrics

Question 62

A previously healthy 2-year-old male presents with sudden onset of coughing, wheezing and difficulty breathing. On examination, he is acutely distressed with respiratory rate of 60/minute, heart rate of 130/min and temperature of 37°C. Lung auscultation reveals monophonic wheezing and decreased breath sounds in the upper zone of the right lung posteriorly.

What is the MOST likely diagnosis?

A. Aspiration pneumonia
B. Acute asthmatic attack
C. Viral wheeze
D. Foreign body aspiration
E. Anaphylaxis

["Knows"]

University of Botswana, Paediatrics
Question 69

A 3-year-old HIV negative boy is seen in the Paediatric General Clinic with a history of easy fatigability that was first noted when he was one year and six months old. His parents report that this complaint is getting worse. His immunization record is up to date.
The picture shown here was taken during the neurological examination of the child.
[The attached images show a child with Duchenne muscular dystrophy]

Which of the diagnoses below is MOST LIKELY to manifest in this way?

A. Chronic inflammatory demyelinating polyneuropathy
B. Poliomyelitis
C. Guillain-Barre dystrophy
D. Duchenne muscular dystrophy
E. Myasthenia gravis

Images showing a child with a specific lower-limb weakness and posture which are classically diagnostic of Duchenne muscular dystrophy were included with Question 69. The question could be answered solely on the basis of the images, and would therefore be testing factual knowledge rather than knowledge application on the basis of the clinical vignette taken together with the images. It was therefore allocated to the "Knows" tier.

In addition, the University of Botswana used **extended matching questions** (EMQs) and **short answer questions** (SAQs) in both disciplines, and these were all recorded at the "Knows How" tier. The examples provided here in Box 5 show a degree of knowledge application and problem solving.

Box 5

["Knows How"]

University of Botswana, Paediatrics
EMQ, Question 111

[Lead-in and Option List]
Poisoning is one of the common reasons for hospital admission in children. Match the following poison situations (options) with the appropriate description (question) below.

A. Alcohol toxicity
B. Salicylic acid (aspirin)
C. Cocaine toxicity
D. Iron toxicity
E. Mushroom toxicity
F. Organophosphate toxicity
G. Paraffin toxicity
H. Paracetamol toxicity
I. Scorpion bite

[Question]
A 17-year-old girl is noted to be hyperventilating on admission. Her arterial blood gas shows a mixed high anion gap metabolic acidosis and respiratory alkalosis.

["Knows How"]

University of Botswana, Obstetrics & Gynaecology
SAQ, Question 7

A 21-year-old primigravida lady had her LNMP [Last normal menstrual period] on 17th November 2014. She had an uneventful prenatal follow up until June 1st 2015 when she complained of headache and right upper quadrant abdominal pain. She reported cough with blood stained sputum and shortness of breath for a day but no fever.

Her BP was 170/110 mmHg, pulse rate 108 beats per minute and respiratory rate 30 breaths per minute. Her haemoglobin was 10 g/dL, platelets 45000/µL, AST 150 U/L, ALT 130 U/L and LDH 600 mg/dL.

(a) Identify all her possible diagnoses and complications.
(b) List four additional investigations to support your diagnosis.
(c) Discuss the management of this patient.
(d) List five possible immediate complications that this patient might develop.

Eight out of 10 **short answer questions** (SAQs) (80%) from Makerere University (described as "structured questions" in Obstetrics & Gynaecology) were at the lower "Knows" tier. Questions 2 and 8 illustrated in Box 6 below only require straight factual knowledge recall in order to answer without even referring to the clinical vignette provided.

Box 6

["Knows"]

Makerere University, Obstetrics & Gynaecology
SAQ Question 2

A Gravida 2 Para 1+0 is diagnosed with fetal distress.
(a) Define fetal distress.
(b) Give 5 causes.
(c) Describe management at 9 cm cervical dilatation.

["Knows"]

Makerere University, Obstetrics & Gynaecology
SAQ Question 8

A couple was investigated for infertility. The gynaecologist made a diagnosis of male factor infertility.

(a) What is male factor infertility?
(b) Semen analysis had been performed. What parameters are studies in semen analysis?
(c) Give 5 possible causes of fallopian tube blockage.

In Paediatrics at University of Botswana, one out of the four SAQs (25%) was in the lower "Knows" tier and the remaining three at the "Knows How" tier.

Makerere University was the only study site using **essays** as an assessment method in Obstetrics & Gynaecology (See Box 7).

Box 7

["Knows How"]

Makerere University, Obstetrics & Gynaecology
Essay, Question 1

Discuss the six pillars of safe motherhood.

As discussed earlier in Chapter 2 Table 2.3, essays are open-ended questions in which the student can provide their answers usually without any limit on the length or number of words (Wass et al., 2001c; Schuwirth and van der Vleuten, 2003a; Schuwirth and van der Vleuten, 2004). They can be structured or unstructured, and in the case of the example given above, this question on the Six Pillars of Safe Motherhood is

unstructured. Essays are time consuming to answer and can only test a very limited number of issues in the time allocated for testing. They are also time consuming to mark, and difficult to mark consistently thereby contributing to their limited reliability, as discussed previously in Section 2.5. However, the question topic in the example given is highly relevant to the context of health-care in Uganda and the Eastern Sub-Saharan African region, so it would be reasonable to expect students to apply their knowledge on this topic and to discuss the significance of safe motherhood while referencing the context in which they will be practising. This would indicate it belonged to the "Knows How" tier of Miller's 'pyramid'. However, it could be argued that the students could equally simply regurgitate the memorised contents of a lecture on safe motherhood without any application of this knowledge, in which case the question would be relegated to the "Knows" tier. Interpretation of the word "Discuss" in this question appears to be key, as well as the answer expected from the students. I was, however, not able to gain access to the model answers for the written examination papers in this instance. Therefore, based on my own clinical judgement and past experience, after consultation with my study supervisor, and for the purposes of analysis of content using Miller's 'pyramid' I determined that this question belonged to the "Knows How" tier. The issues raised here will be discussed further in Chapter 6.

Turning to the **performance-based assessment** methods, the **OSCE** questions extended from "Knows" upwards to the "Knows How" and "Shows How" tiers. With regards to the "Shows How" tier, although one would expect OSCEs to feature strongly at this level, closer review of the station questions revealed that not all stations had a "Shows How" component requiring demonstration of some clinical skill(s) and/or were interactive.

An example here in Box 8 is drawn from the Makerere University Paediatrics OSCE:

Box 8

["Knows"]

Makerere University, Paediatrics
Station 9

[Equipment provided at this OSCE station: Nebuliser device with mask]

(a) Name the specimen and its major constituent parts.
(b) List the clinical indications that require the use of this specimen.
(c) Describe how this specimen is used on the wards.

This OSCE station did not have an examiner present. Students were required to provide a written answer to the questions. The questions relate to recognition of the device provided and require knowledge recall in order to answer. Therefore, I allocated this question to the "Knows" tier of the 'pyramid'. It is worth noting that although this question was situated in an OSCE examination, the format of this station more closely resembled a written SAQ-type question than an OSCE (Refer back to Table 2.3) and did not make optimal use of OSCE as an assessment method.

A similar occurrence took place in the example provided here in Box 9, drawn from the Stellenbosch University Obstetrics & Gynaecology **OSCE/OSPE**, Station 6:

Box 9

["Knows"]

Stellenbosch University, Obstetrics & Gynaecology
Station 6 (on Day 2 of OSCE/OSPE)

A 64-year-old woman, Gravida 4 Para 3 Miscarriage 1, is referred to you by the clinic sister with a complaint of urinary incontinence.

(a) What types of urinary incontinence in females do you know?
(b) Which single most important feature would be characteristic of each of these types?

She leaks urine when she coughs or laughs. She also has the sensation of urgency and at times leaks urine before she gets to the toilet.
(c) What is your diagnosis?
(d) How would you treat her?

This OSCE/OSPE was conducted over two consecutive days, with a separate examination paper for each day. The OSCE station (in Box 9) also did not have any examiner present and students were required to provide written answers to the questions by recalling factual knowledge. The short clinical vignette did not really provide any additional information that contributed towards the answers. The use of clinical vignettes is a common practice in medical examinations in order to try and test a student's problem-solving skills in relation to the particular clinical context described (Case and Swanson, 2002). For example, the management of a child presenting with fever may differ depending on the urgency of the clinical presentation (such as a child with fever who appears relatively well compared to a child with fever who has the

altered level of consciousness of irritability and who has been vomiting). In the above example of Station 6 from Stellenbosch University Obstetrics & Gynaecology, it would be possible to just ask the four questions without the clinical vignette provided, as the student is being tested on knowledge recall of urinary incontinence. It would seem that testing at this lower level does not provide sufficient evidence to help make decisions regarding the clinical competence of medical graduates in relation to local health care needs and raises questions about whether the content of assessment is valid.

In the Stellenbosch University Obstetrics & Gynaecology "OSPE/OSCE", the four "OSPE" stations facilitated by examiners were at the "Knows How" tier. One example is described here in Box 10:

Box 10

["Knows How"]

Stellenbosch University, Obstetrics & Gynaecology
Station 9/10 (on Day 1 of OSCE/OSPE)

The patient is 32 years old and Para 3 Gravida 4. Immediately after vaginal delivery of her baby (singleton), she starts bleeding actively and becomes clinical shocked.

Prepare yourself to discuss the management of the case at the following station.

There was no patient or mannequin (model) and no equipment provided at this station. There was an examiner present and the interaction was purely of an oral nature. Although the question relates to a relevant clinical situation of postpartum haemorrhage and could lend itself to the demonstration of skills related to the management of acute haemorrhage (performed on a mannequin), students were not required to actually "Show How" for any aspect of the clinical management in question. As the station was more of a theoretical nature only requiring verbal responses from the students (or it could also equally well have been assessed in written format), it tested only to the level of "Knows How". The use of an examiner in this way would seem to be a waste of precious and limited human resources, especially in this Sub-Saharan African context.

Only three of the eight "OSCE" stations in the Stellenbosch University Obstetrics & Gynaecology "OSPE/OSCE" with clinical problems were at the "Shows How" tier and

were facilitated by an examiner, and of the remaining five stations that were unmanned and related to data interpretation or the management of clinical problems, four were at the "Knows How" and one at the "Knows" tier. A "Shows How" question is illustrated as follows in Box 11:

Box 11

["Shows How"]

Stellenbosch University, Obstetrics & Gynaecology
Station 1 (on Day 1 of OSCE/OSPE)

This primigravida did not receive any antenatal care and now presents with complaints of severe headache, a blood pressure of 150/110 mmHg and +++ proteinuria. She cannot remember the date of her last menstrual period.

Demonstrate how you would perform the abdominal examination and mention all your findings (positive and negative).
[Mannequin of a pregnant torso provided at the station]

There were similar findings at Makerere University. In Obstetrics & Gynaecology, four out of six OSCE stations were at "Shows How" tier, and the remaining two at "Knows How" tier (See Table 5.8). The 12-station OSCE used in Paediatrics revealed seven, two and three stations at "Shows How", "Knows How" and "Knows" tiers respectively (See Table 5.9). The five stations at the lower tiers of Miller's 'pyramid' were written stations that did not have examiners present. Examples of questions that were allocated to the "Shows How" and "Knows How" tiers of the 'pyramid' are illustrated in Boxes 12 and 13 respectively (an example at the "Knows" tier has already been provided earlier in Box 8). The "Shows How" OSCE stations here were conducted using real patients and required the demonstration of physical examination as well as aspects of clinical management.

Box 12

["Shows How"]

Makerere University, Paediatrics
Station 5

You are asked to review this child who has been unwell for some time now.

(a) Carry out a relevant general examination of this child. Describe what you are doing as you perform the exam.

(b) Summarise the significant findings from the exam.
(c) List (up to 3) important problems you have found in this child.
(d) Outline 3 important investigations you will do in the management of this child.

["Shows How"]

Makerere University, Obstetrics & Gynaecology
Station 6

This is a 32-year-old Para 1+0 admitted with vague lower abdominal pain for 2 months.

(a) Carry out an abdominal examination on this patient.
(b) What other examination would you like to perform?
(c) Give 4 differential diagnoses.

The "Knows How" OSCE stations required the student to interpret data provided. Although (as in the examples illustrated here in Box 13) real patient radiograph and partogram exhibits were used for these questions, there were no patients at the stations. Similar to some of the OSCE/OSPE stations in Obstetrics & Gynaecology at Stellenbosch University, the interactions between examiner and student at these stations were of an oral nature, with the student only being tested on application of knowledge.

Box 13

["Knows How"]

Makerere University, Paediatrics
Station 10

This is a radiograph of one of the patients on the ward.

(a) Read this X-ray.
(b) Outline the important findings from the X-ray.
(c) List 3 likely differential diagnoses. Give reasons to support each of your answers.

["Knows How"]

Makerere University, Obstetrics & Gynaecology
Station 2

NK , a Gravida 3 Para 2+0, is in labour and she is being monitored using the partograph provided.

(a) Identify 4 problems that are depicted in the plotted partograph.

(b) Given the above findings and that further examination show that she has dry lips and tongue, outline the management plan.
(c) Give 2 possible maternal complications.
(d) Give 2 possible fetal complications.

In contrast, all OSCE stations in the medical-based and surgical-based OSCEs at University of Botswana were at the "Shows How" tier. For example (see Box 14):

Box 14

["Shows How"]

University of Botswana, Paediatrics
Station 2

This is the mother of an 8-day-old baby who has just died of septicaemia in the Neonatal Unit at Princess Marina Hospital. The baby's blood culture grew *Klebsiella pneumoniae* sensitive to meropenem (which baby had received appropriately).

The mother does not know her baby has died. Inform her about the sad news.

As illustrated in Tables 5.8 and 5.9, there were no questions in any of the OSCEs, or for that matter, in any of the other assessment methods used, that fell into the "Does" tier. Workplace-based assessment of students in the actual clinical setting where they would normally be practising would be categorised in this tier, but this method was not used as exit-level assessment at the three study sites. Referring back to Tables 5.2 and 5.3 and the earlier discussion on assessment taking place during the clinical clerkships (Section 5.2.1.2), it appears that some of the in-course assessment might be where some assessment at the "Does" tier could be anticipated, for example, with the use of logbooks, mini Clinical Evaluation Exercises (mini-CEX), procedures and Directly Observed Procedural Skills (DOPS).

The other performance-based assessment method used was the **long cases** at Stellenbosch University in Paediatrics and at Makerere University in Obstetrics & Gynaecology. Students were observed interacting with a real patient in a clinical-ward setting and tested on aspects of history taking, physical examination, making a differential diagnosis on the basis of findings, discussion of investigations and clinical management options. Based on these activities which take place during long cases, they were categorised as "Shows How" and not as "Does". As described in Section 4.6.2

outlining the coding framework used to allocate examination questions to different tiers of Miller's 'pyramid', this does not meet the criteria outlined for "Does". The long cases were therefore allocated to the "Shows How" category.

From the analysis presented thus far, based on document analysis and observation in final examinations, the assessment methods described and used seem to focus more on knowledge recall and application (and therefore corresponding to the "Knows" and "Knows How" tiers of Miller's 'pyramid') than on providing students with the opportunity to demonstrate their clinical competence in a tangible manner. As illustrated in Tables 5.8 and 5.9, University of Botswana had the highest percentage of exit-level assessment questions at the higher tiers of Miller's 'pyramid', followed by Stellenbosch University, and with Makerere University remaining at the lowest "Knows" and "Knows How" tiers. This clearly has relevance in terms of making a judgement as to the appropriateness of the evidence that exit-level assessment is generating, and the implications of this will be explored in Chapter 6.

5.3.2 Analysis of content validity using Disability Adjusted Life Years (DALYs)

The second of the two tools used for the analysis looked into the content validity of exit-level assessments in the three study sites. Disability Adjusted Life Years (DALYs) for disease conditions and injuries have been used as indicators of regional health issues in Global Burden of Disease studies, as already described in Chapter 3. In an attempt to appraise the alignment between the education of health-care professionals and local health-care issues pertinent to a region, the relevance of content included in exit-level assessment was explored using DALYs as an analytical framework.

Some general points are noted about the tables and data analysis that are to follow, and these will be mentioned here first. Overall, the 20 leading causes of DALYs for the respective Sub-Saharan African regions were not all relevant to the disciplines of Obstetrics & Gynaecology and Paediatrics (for example, chronic obstructive pulmonary disease, cerebrovascular disease, major depressive disorder, ischaemic heart disease). There were several causes of DALYs that were common to both disciplines, and these included communicable diseases such as HIV/AIDS, tuberculosis and malaria, and there were some causes of DALYs relevant to either one or both disciplines but which fell outside the 20-leading causes (for example, sickle cell disorders, syphilis, poisoning). Paediatric-related causes of DALYs (including lower-respiratory tract infections,

diarrhoeal disease, preterm-birth complications, neonatal encephalopathy, sepsis and other infectious disorders of the newborn baby, protein-energy malnutrition and congenital anomalies) tended to be more spread out across the DALYs tables. "Maternal disorders", although classified as a single cause of DALYs, actually encompassed a number of disorders arising during pregnancy. As a result, there tended to be a concentration of examination question topics around this cause in all three study sites.

The following tables summarise the selection of examination question topics in the final examinations of the three study institutions to causes of DALYs in the corresponding Sub-Saharan African regions:

Table 5.10	Stellenbosch University	Obstetrics & Gynaecology
Table 5.11	Makerere University	Obstetrics & Gynaecology
Table 5.12	University of Botswana	Obstetrics & Gynaecology
Table 5.13	Stellenbosch University	Paediatrics
Table 5.14	Makerere University	Paediatrics
Table 5.15	University of Botswana	Paediatrics

Some question topics emerged below the top-20 DALY ranking, but these were also included to provide an indication of their position in the ranking of causes of DALYs. Tables 5.10 to 5.12 are grouped together, as they relate to maternal-health issues and Tables 5.13 to 5.15 form a second group, relating to issues around child health, in order to provide an overall picture and for ease of comparison of data between the three study sites for the two disciplines.

Furthermore, examination question topics that did not feature in the above tables are also listed separately as follows:

Table 5.10a	Stellenbosch University	Obstetrics & Gynaecology
Table 5.11a	Makerere University	Obstetrics & Gynaecology
Table 5.12a	University of Botswana	Obstetrics & Gynaecology
Table 5.13a	Stellenbosch University	Paediatrics
Table 5.14a	Makerere University	Paediatrics
Table 5.15a	University of Botswana	Paediatrics

Although not the focus of this study, the data is provided to indicate the scope of topics covered in the exit-level assessments and also to provide some context for comparison in terms of content validity. Of note, question topics that related to issues of health promotion and disease prevention, or of "normal" health-care, did not feature in the tables as these were not disease conditions and would therefore not have been regarded as contributory causes of DALYs. Some question topics did not feature high on the list, such as malignancies (cancers of the ovary, uterus and cervix), haematological disorders (including leukaemia), and infertility. (For the overall record of examination questions topics listed for each assessment method in the two disciplines studies, refer to Appendices Q for Stellenbosch University, U for University of Botswana, and V for Makerere University). These limitations of DALYs as an analytical framework that emerged in the course of data analysis will be discussed in Section 6.5.

The discussion will now turn to the findings in the tables, starting with the focus on Obstetrics & Gynaecology.

5.3.2.1 Obstetric & Gynaecological Issues

The causes of DALYs in the Southern Sub-Saharan African (SSA) region are tabulated in Table 5.10. The leading cause for this region was HIV/AIDS, and this topic featured in only one examination question in the Stellenbosch University Obstetrics & Gynaecology OSCE/OSPE examination. The DALYs ranking for maternal disorders in the Southern SSA region was 28. Three OSCE/OSPE question topics fit into the maternal disorders group, but a larger number of question topics related to gynaecological diseases, situated at the lower end of Table 5.10. The yellow shading in this lower portion of the table indicates that there were examination question topics relating to causes of DALYs that did not have a ranking in this World Health Organisation (WHO) region.

The OSCE/OSPE examination at Stellenbosch University was conducted over two consecutive days, with a separate examination paper set for each day. Although equal numbers of obstetrics and gynaecology question topics were set in each examination paper, half the examination question topics did not feature in Table 5.10 but were listed in Table 5.10a instead, as the topics covered issues that did not fall into the listed causes of DALYs. Examples included physical examination of systems, data interpretation of clinical charts, and contraception. These topics concern issues of normal health-care,

health promotion and disease prevention, and would therefore not really be associated with the burden of disease. This will be discussed further in the following Chapter 6.

Table 5.10 Relation of question topics and exit-level assessment methods in the Obstetrics & Gynaecology Final Examinations at Stellenbosch University to causes of Disability Adjusted Life Years (DALYs) in South Africa for females aged 15 to 49 years

% of total DALYs (Female 15-49 years)	Ranking of DALYs in the Southern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in SOUTH AFRICA	Performance-Based	
			OSCE/OSPE * (12)	
			(Day 1*)	(Day 2*)
62.72	**1**	HIV/AIDS	• HIV in pregnancy	
1.34	**2**	Lower respiratory infections		
0.74	**3**	Diarrhoeal disease		
3.77	**4**	Tuberculosis		
	5	Interpersonal violence		
	6	Preterm birth complications		
	8	Diabetes mellitus		
0.87	**11**	Iron deficiency anaemia		
	12	Neonatal encephalopathy (includes birth asphyxia / trauma)		
	13	Road injury		
0.82	**16**	Exposure to mechanical forces		
0.21	**17**	Congenital anomalies		
0.08	**20**	Malaria		
0.13 **0.12** **0.22** **0.12** **0.20**	**28**	Maternal disorders (Includes: - Complications of abortion - Maternal haemorrhage - Maternal hypertensive disorders - Maternal sepsis and other maternal infection - Obstructed labour	• Postpartum haemorrhage	• Antepartum haemorrhage at 32 weeks' gestation

% of total DALYs (Female 15-49 years)	Ranking of DALYs in the Southern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in SOUTH AFRICA	Performance-Based	
			OSCE/OSPE * (12)	
			(Day 1*)	(Day 2*)
0.19 **0.26** **0.18** **0.14**		- Late maternal deaths - Indirect maternal deaths - Maternal deaths aggravated by HIV/AIDS) - Other maternal disorders		• Abdominal pain at 32 weeks' gestation
0.76		Gynaecological diseases	• Vaginal infections (3 wet mount slides) • Adnexal cystic mass on routine gynaecological examination • Vaginal bleeding with 7 weeks' amenorrhoea	• Vaginal bleeding (Ectopic pregnancy) • Urinary incontinence • Pelvic infection • Menorrhagia and dysmenorrhoea
0.25		Cancer of the breast		
0.24		Cancer of the cervix		
0.05		Cancer of the ovary		
0.0089		Cancer of the uterus		• Ca endometrium

Abbreviations: OSCE = Objective Structured Clinical Examination
OSPE = Objective Structured Practical Examination

Explanatory Notes:
Figures in parentheses refer to the total number of questions in that assessment method
• Each bullet point refers to one examination question topic
*OSCE/OSPE conducted over two consecutive days, with a separate examination paper for each day.

• The yellow highlighted section at the bottom of the table refers to examination question topics relating to causes of DALYs (such as gynaecological disease conditions and malignancies) that do not have a ranking in this Southern Sub-Saharan African region.

Table 5.10a Questions topics in the Obstetrics & Gynaecology Final Examinations at Stellenbosch University that could not be related to DALYs and therefore do not feature in Table 5.10

Performance-Based	
OSCE/OSPE * (12)	
(Day 1*)	(Day 2*)
• Abdominal examination (Obstetric) • Pelvic examination • Data interpretation: Antenatal card (Breech presentation) • Data interpretation: Partogram (Poor progress in labour) • Lost intrauterine contraceptive device • Data interpretation: Post-operative observation chart • Patient explanation and counselling (Intrauterine growth retardation)	• Abdominal examination (Obstetric) • Pelvic examination • Data interpretation: Antenatal card (Rhesus iso-immunisation) • Data interpretation: Partogram (Fetal distress) • Episiotomy (types, anatomy, repair)

The data in Table 5.11 relates to Makerere University Obstetrics & Gynaecology examination question topics, which were clearly concentrated under maternal disorders. Makerere University in Uganda is situated in the Eastern SSA region, and therefore the list of causes of DALYs pertains to this region. The DALYs ranking for maternal disorders in the Eastern Sub-Saharan African region was 14. The question topics concerned issues arising during pregnancy, including haemorrhage (whether antepartum or postpartum), infection, pregnancy-induced hypertension, and intrapartum complications (which could also involve consequences for the baby, such as cord prolapse and fetal distress). These topics were spread across a range of assessment methods, both written and performance-based.

Topics other than maternal disorders that featured in this table included HIV (ranked first as a cause of DALYs in Uganda) as well as diabetes mellitus in pregnancy, sickle cell disease, gynaecological diseases and malignancies. Sickle cell disease is prevalent in the Eastern SSA region and therefore examination questions on this topic would be relevant. This lends weight to the use of DALYs as an analytical tool for content validity.

A comparison of Tables 5.10 and 5.11 provide a stark illustration of the advantages as regards content validity that the use of short written formats such as MCQs and SAQs offer over performance-based assessment.

Similarly to the findings from Stellenbosch University (as indicated in Table 5.10a), a number of topics covered issues that did not fall into the listed causes of DALYs and these are shown in Table 5.11a.

Table 5.11 Relation of question topics and exit-level assessment methods in the Obstetrics & Gynaecology Final Examinations at Makerere University to causes of Disability Adjusted Life Years (DALYs) in Uganda for females aged 15 to 49 years

% of total DALYs (Female 15-49 years)	Ranking of DALYs in the Eastern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in UGANDA	Written			Performance-Based	
			MCQ (14)	SAQ (10)	Essay (2)	OSCE (6)	Long Case (1)
30.98	1	HIV/AIDS	• HIV				
4.19	2	Malaria					
2.38	3	Lower respiratory infections					
2.27	4	Diarrhoeal disease					
	5	Preterm birth complications					
	6	Protein energy malnutrition					
5.56	7	Tuberculosis					
	8	Sepsis and other infectious disorders of the newborn baby					
	9	Neonatal encephalopathy (includes birth asphyxia / trauma)		• Umbilical cord prolapse • Fetal distress			
1.40	10	Meningitis					
	11	Road injury					
1.48	12	Iron deficiency anaemia					
	13	Major depressive disorder					

% of total DALYs (Female 15-49 years)	Ranking of DALYs in the Eastern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in UGANDA	Written			Performance-Based	
			MCQ (14)	SAQ (10)	Essay (2)	OSCE (6)	Long Case (1)
1.7 **1.04** **0.96** **0.88** **0.66** **0.83** **0.71** **0.25** **2.34**	**14**	Maternal disorders (Includes: - Complications of abortion - Maternal haemorrhage - Maternal hypertensive disorders - Maternal sepsis and other maternal infection - Obstructed labour - Late maternal deaths - Indirect maternal deaths - Maternal deaths aggravated by HIV/AIDS) - Other maternal disorders	• Post-partum haemorrhage • Pregnancy-induced hypertension • Forceps delivery • Deep venous thrombosis in pregnancy	• Incomplete septic abortion • Antepartum haemorrhage (placenta praevia) • Pregnancy-induced hypertension • Delay in 2nd stage of labour	• Six pillars of safe motherhood	• Post-partum haemorrhage • Delay in 2nd stage of labour	Examples seen: • Fever at term • Vaginal bleeding in 1st trimester
	15	Syphilis					
	29	Diabetes mellitus	• Diabetes mellitus in pregnancy				
	67	Sickle cell disorders	• Sickle cell disease in pregnancy				
0.86		Gynaecological diseases		• Infertility			Example seen: Uterine fibroids
0.55		Cancer of the breast					
0.91		Cancer of the cervix			• Cancer of the cervix		

% of total DALYs (Female 15-49 years)	Ranking of DALYs in the Eastern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in UGANDA	Written			Performance-Based	
			MCQ (14)	SAQ (10)	Essay (2)	OSCE (6)	Long Case (1)
					• Strategies in prevention of cervical cancer		
0.23		Cancer of the ovary	• Cancer of the ovary	• Cancer of the ovary			
0.028		Cancer of the uterus		• Cancer of the endometrium			

Abbreviations: MCQ = Multiple Choice Questions SAQ = Short Answer Questions OSCE = Objective Structured Clinical Examination	**Explanatory Notes:** Figures in parentheses refer to the total number of questions in that assessment method • Each bullet point refers to one examination question topic (unless stated otherwise) • The yellow highlighted section at the bottom of the table refers to examination question topics relating to causes of DALYs (such as gynaecological disease conditions and malignancies) that do not have a ranking in this Eastern Sub-Saharan African region.

Table 5.11a Questions topics in the Obstetrics & Gynaecology Final Examinations at Makerere University that could not be related to DALYs and therefore do not feature in Table 5.11

Written			Performance-Based	
MCQ (14)	**SAQ (10)**	**Essay (2)**	**OSCE (6)**	**Long Case (1)**
• Contraception: Hormonal implant • Contraception: Intrauterine device • Cancer of the vulva • 1st trimester abortion • Normal menstrual cycle • Rhesus isoimmunisation	• Gestational trophoblastic disease		• Abdominal examination (Obstetric) • Abdominal examination (Gynaecological) • Pelvic examination • Contraception	

In Table 5.12, it is evident that a greater number of obstetrics and gynaecology question topics from the University of Botswana could be related to DALYs than was the case for Stellenbosch and Makerere Universities. The University of Botswana obstetric examination question topics were also concentrated under maternal disorders; however the DALYs ranking here was 28. In comparison to the topics at Makerere University, similar issues arising during pregnancy (including haemorrhage, infection, consequences of pregnancy-induced hypertension and intrapartum complications) featured in the University of Botswana final examinations and were spread across a range of assessment methods. Examination question topics on gynaecological disorders (such as urinary incontinence, ectopic pregnancy, pelvic infection and malignancies) were listed, a feature common to all three study sites.

Another common feature of all three study sites was that performance-based assessment provided an opportunity to test physical-examination skills that are relevant to the discipline Obstetrics & Gynaecology. The ability to perform both abdominal and pelvic examinations is important and could contribute to decisions on the clinical competence of medical graduates. However, as these question topics were not related to DALYs, they did not appear in the DALYs tables but instead were listed in Tables 5.10a, 5.11a and 5.12a.

The discussion will next turn to findings in Paediatrics at all three study sites.

Table 5.12 Relation of question topics and exit-level assessment methods in the Obstetrics & Gynaecology Final Examinations at University of Botswana to causes of Disability Adjusted Life Years (DALYs) in Botswana for females aged 15 to 49 years

% of total DALYs (Female 15-49 years)	Ranking of DALYs in the Southern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in BOTSWANA	Written			Performance-Based
			MCQ (12)	EMQ (5)	SAQ (5)	OSCE (4)
46.25	1	HIV/AIDS			• HIV in pregnancy & breastfeeding	
3.68	2	Lower respiratory infections				
2.11	3	Diarrhoeal disease				
6.58	4	Tuberculosis				
	5	Interpersonal violence			• Sexual assault	
	6	Preterm birth complications				
	8	Diabetes mellitus	• Complications of diabetes in pregnancy			
1.06	11	Iron deficiency anaemia				
	12	Neonatal encephalopathy (includes birth asphyxia / trauma)				
	13	Road injury				
0.28	17	Congenital anomalies	• Congenital hormonal abnormality			
0.44	20	Malaria				
0.78 0.55 0.71	28	Maternal disorders (Includes: - Complications of abortion - Maternal haemorrhage - Maternal hypertensive disorders	• Bleeding 2 weeks post-partum	• Antepartum haemorrhage • Eclampsia Q	• Eclampsia	• Incomplete septic abortion

% of total DALYs (Female 15-49 years)	Ranking of DALYs in the Southern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in BOTSWANA	Written			Performance-Based
			MCQ (12)	EMQ (5)	SAQ (5)	OSCE (4)
0.48 0.41 0.58 0.56 0.54 0.43		- Maternal sepsis and other maternal infection - Obstructed labour - Late maternal deaths - Indirect maternal deaths - Maternal deaths aggravated by HIV/AIDS) - Other maternal disorders	• Premature rupture of membranes • Mitral valve lesion in pregnancy	• Trial of labour	• Management of post-dates	
	41	Syphilis		• Syphilis		
0.76		Gynaecological diseases	• Management of infertility • Asherman's syndrome • Ectopic pregnancy • Menorrhagia due to myomata		• Complete prolapse of uterus	
0.19		Cancer of the breast				
0.19		Cancer of the cervix		• Cervical intraepithelial neoplasia (due to Human Papilloma Virus)		
0.04		Cancer of the ovary				
0.0076		Cancer of the uterus				

Abbreviations:	MCQ = Multiple Choice Questions EMQ = Extended Matching Questions SAQ = Short Answer Questions OSCE = Objective Structured Clinical Examination	**Explanatory Notes:** Figures in parentheses refer to the total number of questions in that assessment method • Each bullet point refers to one examination question topic • The yellow highlighted section at the bottom of the table refers to examination question topics relating to causes of DALYs (such as gynaecological disease conditions and malignancies) that do not have a ranking in this Southern Sub-Saharan African region.

Table 5.12a Questions topics in the Obstetrics & Gynaecology Final Examinations at University of Botswana that could not be related to DALYs and therefore do not feature in Table 5.12

Written			Performance-Based
MCQ (12)	**EMQ (5)**	**SAQ (5)**	**OSCE (4)**
• Molar pregnancy • Contraception • Post-operative urinary catheterisation			• Abdominal examination (Gynaecological) • Abdominal examination (Obstetric) • Pelvic examination and Papanicolou smear

5.3.2.2 Paediatric Issues

The causes and rankings of DALYs used for both Stellenbosch University and the University of Botswana are the same since South Africa and Botswana are both situated in the Southern SSA region. Some findings that pertain to both study sites are therefore discussed together here. In Tables 5.13 and 5.15, for Stellenbosch University and the University of Botswana respectively, the leading causes of DALYs which have relevance to children under-five years of age and neonates in the Southern SSA region were lower-respiratory infections (DALYs ranking number 2), diarrhoeal disease (number 3), preterm-birth complications (number 6), neonatal encephalopathy (number 12), and congenital anomalies (number 17). These topics were spread across a range of assessment methods used at Stellenbosch University and the University of Botswana, both in written and performance-based formats.

It is worth noting that in the "slide written" examination at Stellenbosch University, 13 out of 19 question topics feature in the table of causes of DALYs (Table 5.13).

Attention is also drawn to the Paediatrics performance-based assessment at the University of Botswana which needs a little further explanation. Two of the three OSCE stations listed in Table 5.15 required the demonstration of skills. The station with the question topic on Down's syndrome featured a live patient (child) with this condition. Students were required to demonstrate physical examination of a child with dysmorphic features in order to arrive at a diagnosis. A second OSCE station with the

question topic of septicaemia was to test an aspect of communication skills. Students were required to break the sad news of a baby's death from septicaemia (caused by the *Klebsiella pneumoniae* organism) to the baby's mother (role-played by a standardised patient). Another two OSCE stations with question topics that could not be related to DALYs (the demonstration of examination of the cardiovascular and abdominal systems) are listed in Table 5.15a.

Table 5.13 **Relation of question topics and exit-level assessment methods in the Paediatrics Final Examinations at Stellenbosch University to causes of Disability Adjusted Life Years (DALYs) in South Africa for children aged under-5 years, and 5-14 years**

% of total DALYs (Children under 5 years)	% of total DALYs (Children 5-14 years)	Ranking of DALYs in the Southern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in SOUTH AFRICA	Written	Performance-Based
				"Slide Written" (19)	Long cases (2)
5.97	41.04	1	HIV/AIDS	• HIV positive mother; Infant feeding practices	
13.98	1.64	2	Lower respiratory infections	• Acute lower respiratory tract infection	
15.79	1.91	3	Diarrhoeal disease	• Metabolic acidosis in severe diarrhoea and vomiting	
1.02	1.12	4	Tuberculosis	• TB meningitis	Clinical cases seen: • TB meningitis • Pulmonary TB
	0.71	5	Interpersonal violence		
10.32	0.29	6	Preterm birth complications	• Premature baby with apnoea	
		8	Diabetes mellitus	• Diabetic lipohypertrophy	
1.70	8.40	11	Iron deficiency anaemia		
7.73	0.35	12	Neonatal encephalopathy (includes birth asphyxia / trauma)	• Cerebral palsy / Duchenne muscular dystrophy	Clinical case seen: • Cerebral palsy with spastic quadriplegia, failure to thrive, HIV positive
	1.76	13	Road injury		
	0.84	16	Exposure to mechanical forces		
5.28	1.28	17	Congenital anomalies	• Trisomy 21 (Down's syndrome) • Congenital adrenal hyperplasia • Biliary atresia • Tetralogy of Fallot	Clinical case seen: • Patent ductus arteriosus
		19	Epilepsy		
0.056	0.23	20	Malaria		

% of total DALYs (Children under 5 years)	% of total DALYs (Children 5-14 years)	Ranking of DALYs in the Southern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in SOUTH AFRICA	Written	Performance-Based
				"Slide Written" (19)	Long cases (2)
		25	Chronic kidney diseases	• Nephrotic syndrome	
4.16		**29**	Sepsis and other infectious disorders of the newborn		
3.45	**0.17**	**36**	Protein energy malnutrition	• Protein energy malnutrition	

Explanatory Notes:
Figures in parentheses refer to the total number of questions in that assessment method
• Each bullet point refers to one examination question topic

Table 5.13a Questions topics in the Paediatrics Final Examinations at Stellenbosch University that could not be related to DALYs and therefore do not feature in Table 5.13

Written	Performance-Based
"Slide Written" (19)	**Long cases (2)**
• Asthma, atopy • Thrombocytopenia / Idiopathic thrombocytopenic purpura • Neurofibromatosis • Cystic fibrosis • Data interpretation: Electrocardiogram + Resuscitation of infant • Data interpretation: Scarlet fever rash	

There are some similarities in the rankings of causes of DALYs between the Eastern and Southern SSA regions. For example, HIV/AIDS, lower-respiratory infections, diarrhoeal disease, preterm-birth complications, and neonatal encephalopathy are found to be some of the leading causes of DALYs in both SSA regions. In contrast, there are also regional differences which perhaps highlight health-care issues that have significance in that particular region. For instance, malaria has a DALYs ranking of 2 in the Eastern SSA region (Table 5.14), compared to number 20 in the Southern SSA region Tables 5.13 and 5.15. Protein-energy malnutrition (PEM) ranks as number 6 in the Eastern SSA region (Table 5.14), but only features as number 36 in the Southern SSA region (Table 5.13).

In the Makerere University Paediatrics final examinations, two-thirds of the 50 MCQs recorded examination question topics that related to relevant paediatric causes of DALYs for this SSA region. One of the four short answer questions (SAQs) related to diarrhoeal diseases (Table 5.14, DALYs ranking number 4) and protein-energy malnutrition (number 6). Three out of 12 OSCE question topics also related to relevant paediatric causes of DALYs: HIV/AIDS (number 1), sepsis in the newborn (number 8), and cardiomyopathy (number 44).

Some conditions may be more pertinent to a particular SSA region as they occur more frequently, such as poisonings and sickle cell disorders, which appeared in MCQs in the Makerere University Paediatrics final examinations (see Table 5.14, DALYs rankings number 40 and 67 respectively), although poisonings also featured in extended matching questions (EMQs) in the University of Botswana Paediatrics final

examinations (see Table 5.15, DALYs ranking number 64). A greater number of paediatrics question topics from Stellenbosch University could be aligned with to DALYs than was the case for Makerere University and the University of Botswana.

Table 5.14 **Relation of question topics and exit-level assessment methods in the Paediatrics Final Examinations at Makerere University to causes of Disability Adjusted Life Years (DALYs) in Uganda for children aged under-5 years, and 5-14 years**

% of total DALYs (Children under 5 years)	% of total DALYs (Children 5-14 years)	Ranking of DALYs in the Eastern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in UGANDA	Written		Performance-Based
				MCQ (50)	SAQ (4)	OSCE (12)
3.48	14.16	1	HIV/AIDS	• HIV positive pregnant mother • HIV positive child		• HIV (Data interpretation)
17.25	16.4	2	Malaria	• Malaria		
11.68	4.63	3	Lower respiratory infections	• Pneumonia • Pneumonia (treatment) • Bronchiolitis		
7.23	3.72	4	Diarrhoeal disease	• Diarrhoea and vomiting • Severe diarrhoea • Severe diarrhoea (treatment)	• Diarrhoea and weight loss (with protein energy malnutrition)#	
8.38	0.39	5	Preterm birth complications			
5.57	1.77	6	Protein energy malnutrition	• Protein energy malnutrition	• Protein energy malnutrition (with diarrhoea and weight loss)#	
1.08	1.25	7	Tuberculosis	• Pneumonia • BCG vaccination • TB spine		
6.60		8	Sepsis and other infectious disorders of the newborn baby	• Group B streptococcal infection • Group A streptococcal infection		• Ophthalmia neonatorum
8.28	0.23	9	Neonatal encephalopathy (includes birth asphyxia /trauma)	• Neonatal seizures • Respiratory distress syndrome		
3.74	2.83	10	Meningitis	• Meningitis		
	4.68	11	Road injury			
1.16	2.83	12	Iron deficiency anaemia			
		14	Maternal disorders			

% of total DALYs (Children under 5 years)	% of total DALYs (Children 5-14 years)	Ranking of DALYs in the Eastern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in UGANDA	Written		Performance-Based
				MCQ (50)	SAQ (4)	OSCE (12)
		15	Syphilis			
4.76	**1.81**	**18**	Congenital anomalies	• Congenital hypothyroidism • Down's syndrome • Patent ductus arteriosus		
		19	Epilepsy	• Convulsions		
		20	Chronic obstructive pulmonary disease	• Bronchiectasis		
	0.45	**23**	Interpersonal violence	• Suspected child abuse		
		29	Diabetes mellitus	• Infant of diabetic mother • Diabetes mellitus in children	• Diabetic ketoacidosis	
	0.11	**40**	Poisonings	• Pesticide ingestion • Kerosene ingestion		
		42	Chronic kidney diseases	• Nephrotic syndrome		
		44	Cardiomyopathy and myocarditis	• Rheumatic fever		• Data interpretation(Chest radiograph)
		67	Sickle cell disorders	• Sickle cell anaemia • Sickle cell anaemia (transfusion reaction)		

Abbreviations: MCQ = Multiple Choice Questions
SAQ = Short Answer Questions
OSCE = Objective Structured Clinical Examination

Explanatory Notes:
Figures in parentheses refer to the total number of questions in that assessment method
• Each bullet point refers to one examination question topic (unless stated otherwise)

= Two topics in the same question

Table 5.14a Questions topics in the Paediatrics Final Examinations at Makerere University that could not be related to DALYs and therefore do not feature in Table 5.14

Written		Performance-Based
MCQ (50)	**SAQ (4)**	**OSCE (12)**
• Normal developmental milestones at routine child health check • Differences in blood for newborn and 24-month-old baby • Routine care of newborn at delivery • Physiological jaundice of the newborn • Puberty • Adolescence • Adolescent changes • Jaundice and Hepatitis B • Anaemia (worm infestation) • Leukaemia • Haemophilia • Lactase deficiency • Infant with heart failure • Mild respiratory tract infections • Symptoms of intussusception • Infant with febrile convulsions, encephalitis • Guillain-Barre syndrome • Zinc deficiency	• Neonatal mortality rate in Uganda • Immunisations; government vaccination programme	• Examination: Cardiovascular system • Examination: Respiratory system • Examination: Abdominal system • Infant resuscitation • Examination of an unwell child • Communication: Patient explanation (Blood transfusion for severe anaemia) • Communication: History taking (Neurological system) • Equipment: Nebuliser • Equipment: Mask with rebreather bag

Table 5.15 **Relation of question topics and exit-level assessment methods in the Paediatrics Final Examinations at University of Botswana to causes of Disability Adjusted Life Years (DALYs) in Botswana for children aged under-5 years, and 5-14 years**

% of total DALYs (Children under 5 years)	% of total DALYs (Children 5-14 years)	Ranking of DALYs in the Southern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in BOTSWANA	Written			Performance-Based
				MCQ (12)	EMQ (6)	SAQ (5)	OSCE (5)
2.77	4.47	1	HIV/AIDS			• AIDS	
9.50	2.10	2	Lower respiratory infections	• Pneumococcal infection			
13.47	2.71	3	Diarrhoeal disease	• Acute gastroenteritis			
1.00	1.82	4	Tuberculosis				
	0.53	5	Interpersonal violence				
13.19	2.38	6	Preterm birth complications				
		8	Diabetes mellitus	• Diabetic ketoacidosis		• Baby born to diabetic mother	
1.99	11.55	11	Iron deficiency anaemia				
11.08	0.66	12	Neonatal encephalopathy (includes birth asphyxia / trauma)				
	4.03	13	Road injury				
	1.30	16	Exposure to mechanical forces				
6.88	2.28	17	Congenital anomalies	• Trisomy 21 (Down's syndrome)		• Congenital heart disease	• Down's syndrome • Lower motor neurone lesion / Spinal muscular atrophy
		19	Epilepsy				
1.97	4.61	20	Malaria				
4.61		29	Sepsis and other infectious disorders of the newborn				• Septicaemia (*Klebsiella pneumoniae*)

% of total DALYs (Children under 5 years)	% of total DALYs (Children 5-14 years)	Ranking of DALYs in the Southern Sub-Saharan African Region	Causes of Disability-Adjusted Life Years (DALYs) in BOTSWANA	Written			Performance-Based
				MCQ (12)	EMQ (6)	SAQ (5)	OSCE (5)
	0.13	**64**	Poisoning		• Salicylic acid (aspirin) • Paraffin toxicity • Organophosphate toxicity		

Abbreviations: MCQ = Multiple Choice Questions EMQ = Extended Matching Questions SAQ = Short Answer Questions OSCE = Objective Structured Clinical Examination	**Explanatory Notes:** Figures in parentheses refer to the total number of questions in that assessment method • Each bullet point refers to one examination question topic

Table 5.15a Questions topics in the Paediatrics Final Examinations at University of Botswana that could not be related to DALYs and therefore do not feature in Table 5.15

Written			Performance-Based
MCQ (12)	**EMQ (6)**	**SAQ (5)**	**OSCE (5)**
• Infective endocarditis • Foreign body aspiration • Snoring with enlarged tonsils • Disseminated intravascular coagulation • Factor VIII deficiency • Glasgow Coma Scale • Duchenne muscular dystrophy • Night terror disorder	• Tanner staging of sexual maturity (Scenario 1) • Tanner staging of sexual maturity (Scenario 2) • Tanner staging of sexual maturity (Scenario 3)	• Anaemia (Thalassaemia) • Asthma	• Examination: Cardiovascular system • Examination: Abdominal system

Looking at the two different SSA regions with different rankings of causes of DALYs poses challenges in obtaining an overall view in relation to the scope of examination question topics. For this broader view, one possible way to look at the relevance of the conditions that are causes of DALYs may be to create a "cumulative ranking score". This cumulative ranking score was obtained by adding the ranking from the respective regions to get an impression of the overall importance of the topic (Refer to the three left-hand columns in Table 5.16 which provides a summary of the relation of question topics and exit-level assessment methods to this cumulative ranking score for leading causes of DALYs from the Eastern and Southern SSA regions.). It is evident that examination question topics do cover a spread of the causes of DALYs for the two SSA regions pertinent to this study and, using the cumulative ranking, are mostly situated in the top 50 causes.

One other apparent observation of the table as a whole is that Paediatrics across all three study sites had more examination questions that featured in the DALYs lists than Obstetrics & Gynaecology. If percentages were calculated for the number of questions that featured in DALYs lists in relation to the total number of examination questions in all the assessment methods used in that medical school for that discipline, the figures for Obstetrics & Gynaecology would be 25% at Stellenbosch University, 44% at Makerere University, and 54% at the University of Botswana. In comparison, the figures for Paediatrics would be 68% (Stellenbosch), 5% (Makerere) and 43% (Botswana). These figures provide an indication of the frequency of testing of the question topic, and

together with the cumulative ranking score, would provide suitable evidence in support of determining the content validity of the exit-level assessments.

Table 5.16 Relation of question topics and exit-level assessment methods to the cumulative ranking for leading causes of DALYs from the Eastern and Southern SSA regions

Cumulative Ranking Score	Ranking in Southern SSA region	Ranking in Eastern SSA region	Causes of Disability-Adjusted Life Years (DALYs) in the Eastern & Southern Sub-Saharan African Regions	OBSTETRICS & GYNAECOLOGY			PAEDIATRICS		
				Stellenbosch University	Makerere University	University of Botswana	Stellenbosch University	Makerere University	University of Botswana
2	1	1	HIV/AIDS	1 (OSCE/OSPE)	1 (MCQ)	1 (SAQ)	1 (slide written)	2 (MCQ) 1 (OSCE)	1 (SAQ)
5	2	3	Lower respiratory infections				1 (slide written)	3 (MCQ) 1 (OSCE)	1 (MCQ)
7	3	4	Diarrhoeal disease				1 (slide written)	3 (MCQ) 1 (SAQ)	1 (MCQ)
11	4	7	Tuberculosis				1 (slide written)	3 (MCQ)	
11	6	5	Preterm birth complications				1 (slide written)		
21	12	9	Neonatal encephalopathy (includes birth trauma/asphyxia)	1 (OSCE/OSPE)	2 (SAQ)		1 (slide written)	2 (MCQ)	
22	20	2	Malaria					1 (MCQ)	
23	7	16	Cerebrovascular disease						
23	10	13	Major depressive disorder						
23	11	12	Iron deficiency anaemia					1 (MCQ)	
24	13	11	Road injury						
28	5	23	Interpersonal violence			1 (SAQ)		1 (MCQ)	
29	9	20	Chronic obstructive pulmonary disease					1 (MCQ)	
32	15	17	Low back pain						
34	24	10	Meningitis					1 (MCQ)	
35	14	21	Ischaemic heart disease						

Cumulative Ranking Score	Ranking in Southern SSA region	Ranking in Eastern SSA region	Causes of Disability-Adjusted Life Years (DALYs) in the Eastern & Southern Sub-Saharan African Regions	OBSTETRICS & GYNAECOLOGY			PAEDIATRICS		
				Stellenbosch University	Makerere University	University of Botswana	Stellenbosch University	Makerere University	University of Botswana
35	17	18	Congenital anomalies			1 (MCQ)	4 (slide written)	3 (MCQ)	1 (SAQ) 2 (OSCE)
37	8	29	Diabetes mellitus		1 (MCQ)	1 (MCQ)	1 (slide written)	2 (MCQ) 1 (SAQ)	1 (MCQ) 1 (SAQ)
37	29	8	Sepsis and other infectious disorders of the newborn baby					2 (MCQ) 1 (OSCE)	1 (OSCE)
38	19	19	Epilepsy						
42	28	14	Maternal disorders	1 (OSCE/OSPE)	3 (MCQ) 4 (SAQ) 1 (Essay) 2 (OSCE)	3 (MCQ) 3 (EMQ) 2 (SAQ) 1 (OSCE)			
42	36	6	Protein energy malnutrition				1 (slide written)	1 (MCQ) 1 (SAQ)	
54	16	38	Exposure to mechanical forces						
56	41	15	Syphilis			1 (EMQ)			
57	18	39	Drug use disorders						
67	25	42	Chronic kidney disease				1 (slide written)	1 (MCQ)	
75	31	44	Cardiomyopathy and myocarditis					1 (MCQ)	
104	64	40	Poisoning					2 (MCQ)	3 (EMQ)
201	134	67	Sickle cell disorders					2 (MCQ)	

Abbreviations:
MCQ = Multiple Choice Questions
EMQ = Extended Matching Questions
SAQ = Short Answer Questions
OSCE = Objective Structured Clinical Examination
OSPE = Objective Structured Practical Examination

The question topics and corresponding exit-level assessment methods (that featured previously in Tables 5.10a, 5.11a, 5.12a, 5.13a, 5.14a and 5.15a) that do not feature in Table 5.16 are summarised and presented in Table 5.17 to provide an overall picture of the range of topics. As mentioned previously in Section 5.3.2, some of these question topics are not disease conditions and would therefore not be regarded as contributory causes of DALYs. They relate instead to issues of health promotion and disease prevention, or of "normal" health-care. Examples include abdominal or pelvic examination techniques, normal developmental milestones, immunisations, normal menstrual cycle, puberty and adolescence, staging of sexual maturity and contraception. These topics listed in Table 5.17 are highly relevant to being a doctor in Sub-Saharan Africa. Their inclusion in assessment revealed the limitations of DALYs as a disease-based analytical framework, and this will be discussed further in Section 6.2.1.3.

In summary, an analysis of assessment content has been provided in this Section 5.3. In designing the study much thinking went into what might serve as a suitable proxy for health-related issues in SSA and after careful consideration the decision was made to use the DALYs. In Chapter 6, I reflect on the suitability of DALYs as an analytical tool commenting on both the advantages and possible limitations in using this approach.

Table 5.17 Question topics and corresponding exit-level assessment methods that do not feature in Table 5.16 and have no relation to leading causes of DALYs in the Eastern and Southern SSA regions

OBSTETRICS & GYNAECOLOGY			PAEDIATRICS		
Stellenbosch University	**Makerere University**	**University of Botswana**	**Stellenbosch University**	**Makerere University**	**University of Botswana**
7 (OSCE/OSPE*) *Day 1 **5 (OSCE/OSPE#) #Day 2**	**6 (MCQ)** **1 (SAQ)** **4 (OSCE)**	**4 (MCQ)** **3 (OSCE)**	**6 (slide written)**	**18 (MCQ)** **2 (SAQ)** **9 (OSCE)**	**8 (MCQ)** **3 (EMQ)** **2 (SAQ)** **2 (OSCE)**
• Abdominal examination (Obstetric) • Pelvic examination • Data interpretation: Antenatal card (Breech presentation) • Data interpretation: Partogram (Poor progress in labour) • Lost intrauterine contraceptive device • Data interpretation: Post-operative observation chart • Patient explanation and counselling (Intrauterine growth retardation) • Abdominal examination (Obstetric) • Pelvic examination • Data interpretation: Antenatal card (Rhesus iso-immunisation) • Data interpretation: Partogram (Fetal distress) • Episiotomy (types, anatomy, repair)	• Contraception: Hormonal implant • Contraception: Intrauterine device • Cancer of the vulva • 1st trimester abortion • Normal menstrual cycle • Rhesus Isoimmunisation • Gestational trophoblastic disease • Abdominal examination (Obstetric) • Abdominal examination (Gynaecological) • Pelvic examination • Contraception	• Molar pregnancy • Contraception • Post-operative urinary catheterisation • Abdominal examination (Gynaecological) • Abdominal examination (Obstetric) • Pelvic examination and Papanicolou smear	• Asthma, atopy • Thrombocytopenia / Idiopathic thrombocytopenic purpura • Neurofibromatosis • Cystic fibrosis • Data interpretation: Electrocardiogram + Resuscitation of infant • Data interpretation: Scarlet fever rash	• Normal developmental milestones at routine child health check • Differences in blood for newborn and 24-month-old baby • Routine care of newborn at delivery • Physiological jaundice of the newborn • Puberty • Adolescence • Adolescent changes • Jaundice and Hepatitis B • Anaemia (worm infestation) • Leukaemia • Haemophilia • Lactase deficiency • Infant with heart failure • Mild respiratory tract infections • Symptoms of intussusception • Infant with febrile convulsions, encephalitis • Guillain-Barre syndrome • Zinc deficiency • Neonatal mortality rate in Uganda	• Infective endocarditis • Foreign body aspiration • Snoring with enlarged tonsils • Disseminated intravascular coagulation • Factor VIII deficiency • Glasgow Coma Scale • Duchenne muscular dystrophy • Night terror disorder • Tanner staging of sexual maturity (Scenario 1) • Tanner staging of sexual maturity (Scenario 2) • Tanner staging of sexual maturity (Scenario 3) • Anaemia (Thalassaemia) • Asthma • Examination: Cardiovascular system • Examination: Abdominal system

				• Immunisations; government vaccination programme • Examination: Cardiovascular system • Examination: Respiratory system • Examination: Abdominal system • Infant resuscitation • Examination of an unwell child • Communication: Patient explanation (Blood transfusion for severe anaemia) • Communication: History taking (Neurological system) • Equipment: Nebuliser • Equipment: Mask with rebreather bag	

Abbreviations:
MCQ = Multiple Choice Questions
EMQ = Extended Matching Questions
SAQ = Short Answer Questions
OSCE = Objective Structured Clinical Examination
OSPE = Objective Structured Practical Examination

5.4 Factors influencing assessment practice

The remainder of this chapter relates to the findings from the interviews that took place at the three study sites. The study sites have already been described in some detail in Section 4.3. This section specifically focuses on the third and fourth study objectives ("To explore the reasons for the choices of exit-level assessments" and "To investigate what clinical competencies were considered appropriate for medical graduates from these medical schools in order to address selected health-related issues") and draws on the interviews that were conducted as part of the empirical work within this study.

As described in Section 4.4.3, the interviews were divided into two main groups, involving examiners and interns. As shown in Table 5.18, a total of 23 interviews were conducted at the three study sites with participants being interviewed either in groups or individually. Out of a total of 68 participants, 45 were examiners (19 male, 26 female) and 23 were interns (11 male and 12 female).

The intern interviews did not provide any information that focussed on the third and fourth study objectives and therefore the remainder of this chapter deals with findings from the examiners.

Table 5.18 Overview of the interview participants

		Stellenbosch University	Makerere University	University of Botswana	
Examiners	(10 Group interviews)	14*	19*	4*	No. of participants: 37
	(8 Individual interviews)	3*	2*	3*	No. of participants: 8
Interns	(5 Group interviews)	10*	13*	-	No. of participants: 23
	(Total no. of interviews: 23)				Total no. of participants: 68

Number of participants

The length of involvement of examiners in exit-level assessment in their respective institution (as examiners or examination coordinators) ranged from 1 to 29 years, with the average length of involvement being 8.46 years. Approximately 75% (33

examiners) had been involved in exit-level assessment at their respective institutions as examiners or examination coordinators for less than 10 years' duration. A further breakdown of this number shows that 21 examiners had 0 to 5 years' involvement, and the remaining 12 examiners had 6 to 10 years' involvement. 13% (6 examiners) had been involved between 11 to 20 years, and the remaining 13% (6 examiners) had more than 20 years' experience with assessment. All the examiners interviewed at University of Botswana were in the 0 to 5 year range of involvement in exit-level assessment as this is a young institution.

In exploring reasons for choices of exit-level assessments and their understanding of clinical competence, issues which influenced assessment practice emerged from thematic analysis of the interviews with examiners. These are each described in turn below. Section 5.4.1 addresses aims of the training programme, influences from outside and within the institution are described in Sections 5.4.2 and 5.4.3 respectively, followed by expectations for graduates in Section 5.4.4.

5.4.1 Aim of the training programme

Examiners at University of Botswana emphasized the need to train doctors to be generalists, not specialists, and that this was the overall aim of their medical training programme. The graduates needed to be able to handle a wide variety of clinical conditions in anticipation of the health-care that they would be expected to provide in their context.

> [C02D261:115] "*You know, we're training these doctors not to be a specialist in paediatrics or a specialist in surgery. They're going to be GPs [General Practitioners].*"

Taking cognisance of this overall aim, and following through the argument that patients seen in general-practice context present differently, during the planning stages for the undergraduate medical programme it was felt inappropriate to have separate assessments by discipline and this resulted in integrated examination papers planned for exit-level assessment as was evident at the University of Botswana:

> [C04Y263:113-114] *"The idea of the integrated exam at exit-level was also based on the fact that at this point the emphasis of assessment is to pass out an all-round doctor not so much a specialist in each speciality. Such a doctor is able to pick up, say, a fracture in surgery, evaluate a red*

> *eye in ophthalmology, understand basic anaesthesia etc. There was a lot of discussion and it was finally agreed that having the exam integrated was the better way to go. The challenging question was that if you failed what discipline will you have failed? The consensus is that you have failed to fulfil basic requirements of practising as a general doctor even if you may have passed all questions in one clinical discipline. The mark attained is what you score for all the disciplines. You cannot therefore be passed by knowing a lot in one discipline and hardly anything in another."*

Examiners at Makerere University also talked about "*trying to align competencies with the methods of assessment*" [B04Y120:26] as they realised the importance of ensuring their students had the competencies required for clinical practice. This would be in agreement with Biggs' (1996) constructive alignment of assessment with intended curriculum outcomes, and also speaks to the alignment of education and health-care needs advocated by Frenk et al. (2010), both discussed earlier in Section 3.6. This does emphasize the importance of alignment and could have implications on future curriculum review. Some of these issues will be discussed further in the next chapter.

5.4.2 Influences from outside the institution

Another factor influencing the decisions made with regards to assessment, both in terms of content and approach, related to influences from outside of the institution such as regulatory and licencing bodies. For example, the requirements of the Health Professions Council of South Africa (HPCSA) have to be taken into account in the choices of assessment methods at Stellenbosch University. The HPCSA is the regulatory body for health professions in South Africa and provides guidelines relating to the "registration, education, training, professional conduct and ethical behaviour, ensuring continuing professional development, and fostering compliance with health-care standards" (Health Professions Council of South Africa, 2017).

Peers from other institutions may also have played a part historically in decisions on the choice of assessment methods. As an example, Stellenbosch University was the first in the country to introduce the Objective Structured Clinical Examination (OSCE) circa 1990.

> [A01Y101:175] *"One of our major problems at the time was because the Health Professions Council requires that you pass a clinical examination at the end of your sixth year, is whether they regard this as*

> *clinical….because in those days, people had the frame of mind, clinical means I've got to touch a patient, otherwise it's not clinical…"*

Thus heads of department of other medical schools in South Africa were invited to Stellenbosch University as external examiners. Their positive opinions on the appropriateness of OSCE as a clinical examination helped facilitate its approval by the HPCSA, as it fulfilled the council requirements of a "clinical examination" at exit-level.

In addition collaboration with other institutions beyond the borders of the specific country, notably Europe and North America, resulted in doctors from Sub-Saharan African institutions being sent overseas for specialist training, and those who returned introduced assessment practices based on experiences gained in their respective training sites.

> [B05Y125:84-85] *"…the OSCE came in at a time when, we have collaborations with colleagues in other universities, places like Canada, we have colleagues from the U.S. There is an Association that we have with Obstetricians and Gynaecologists, and we are working with colleagues of ours from Canada. So this idea of the OSCE also came around several years ago at the association level and we looked at its benefits vis-à-vis its challenges, and we thought adopting it would help because it was quite objective."*

While the exchange of ideas and experiences on assessment practices is to be encouraged and supported, there are far-reaching implications arising from their introduction, such as looking into the feasibility of conducting new assessment methods and the training of faculty, whether it is suitable to fit in with the existing curriculum or whether this in turn would require review. What works in one context may not necessarily do so in another. Further discussion will ensue in Chapter 6.

5.4.3 Influences from within the institution

Factors from within the institution, however, proved to be a stronger influence. These included knowledge of institutional requirements.

5.4.3.1 Decision-making with regards to assessment

Decisions were made at the faculty management level, and not at individual department level, during the planning stages for the undergraduate medical programme at a new medical school in Botswana. However, a number of examiners did not appear to know

the reasons for the choice of assessment methods. Examiners' responses suggested that they were not necessarily part of the decision-making process. It is also unclear to what extent examiners were involved in deciding the content of the assessments, as there was no mention of this in any of the interviews.

[C02D261:115] *"…this was not at the department level, this was at the School of Medicine level…"*

Even at an established institution such as Stellenbosch University, it also appeared that faculty were not always informed about the reasons for choices of assessment methods or with changes in the curriculum:

[A03D266:314] *"…when the curriculum changed, it was not communicated to us. We didn't know the details of what was being taught in what years…"*

5.4.3.2 Past practices

Examiners' own experiences of assessment may have played a part in current assessment practices.

[A09D255:082] *"When I was a student….we actually used to have oral exams with a consultant…"*

[B02D114:107] *"Our generation did the long case. Clerk 1 hour, get examined 10-15 minutes."*

[B03Y117:021] *"When I was still a student, we did not have the OSCEs. We had the short case and long case. OSCEs are rather new….."*

[B08D135:028] *"When I joined the university as a student… the clinical exam was the traditional long case and short case, and then vivas…"*

Sometimes decisions were shaped by previous practices, including practical logistics. For example, at Stellenbosch University, long clinical cases were used in Obstetrics and Gynaecology prior to the introduction of the OSCE/OSPE around 1990. It became apparent that with increasing numbers of students in training, a proportionally larger number of clinical cases (live patients) had to be found for the clinical examinations.

[A01Y101:173] *"So [the cases] were becoming less, and also the rule that no patient can be examined twice. So if you have been examined, you're finished. You can't come for ten students."*

Even if an adequate number of clinical cases were available, examiners did not necessarily have control over the circumstances of some of the cases. This was especially so with cases of an obstetric nature. A pregnant woman with a breech presentation or with twins scheduled for the student examinations could have delivered the previous day, necessitating substitute fresh clinical cases to be sought at very short notice.

[A01Y101:173] *"…When you come the next day, you say, I'd like to have those three patients, Oh, they all delivered last night. So, that was becoming a problem. So that was becoming a problem also with gynaecology, because what really happened is we took the patients that were in for an operation tomorrow and say you come for the exam today, they often had an interest, so they were nice patients. So they were becoming less, and also the rule that no patient can be examined twice. So if you've been examined, you're finished, you can't come for ten students."*

Some assessment methods had been in use for some considerable time, so that the original reasons for the choices would not have been apparent to faculty who joined the institution more recently.

[A02D202:68] "*It seems like forever. I tried to ask, but it seems like they have been doing this [slide written test] the same way forever and a day."*

[A08D253:88] "*We just inherited it [laughs]."*

[A08D252:94] "*…it's been like that forever. We have spoken about changing it, and especially now recently last week we spoke about changing the written paper to just multiple choice questions, which we think might be a better reflection of testing more knowledge, and then also easier to mark…"*

[A09D256:94] "*I think for as long as I have worked here it's been like that, so it's been a while. It's probably a good ten years maybe, or more."*

It was interesting to note that when examiners were asked about reasons for choices of assessment methods, many described what assessment methods were used and how they were used in their replies. While this provides a useful source of information to triangulate with the data obtained from document analysis and observations, it does suggest that examiners were not able to provide any explanation or to answer this query directly. Those examiners who were able to provide some explanations for choices of assessment methods tended to be the ones who were senior and experienced

in their respective discipline with more than 20 years' experience with assessment, and who had also been involved at programme-committee level or as part of the senior management in the faculty.

5.4.3.3 Variability in assessment practices and in examiners

Another factor influencing assessment practice related to variability in terms of the actual assessment practices (comprising cases of varying difficulty, the number of clinical cases, assessment question content and the length of examination time) and also variability of examiners.

Examiners at Stellenbosch University commented that the long clinical cases previously used in Obstetrics & Gynaecology were of varying difficulty and, therefore, this in turn could impact on the performance of students.

> [A06Y148:36-37] *"... I was part of the decisions to change towards the OSCE/OSPE type of exam. I think the more important argument at that time which favoured a change was the fact that for an exam to be fair, the candidates must all be exposed to the same level of difficulty of the exam. In other words, if you use clinical cases, quite a proportion of the candidates will of course get quite easy clinical cases because it's easy to use these patients, but there will be a proportion that gets more complicated cases... So that was one very good argument to say, Well, if you use this format, they are all exposed to the same degree of difficulty..."*

Other exit-level assessment practices reflected the result of an evolution of assessment practice over time that mirrored contemporary understanding of good practice. This was the case at Stellenbosch University in Paediatrics with changes to the number of cases that students were examined on. Two short cases with a different examiner each were now used, supplanting the single long case used previously as limitations of the long case became apparent. The format of the end of clerkship examinations was also changed to reduce differences between end of clerkship examinations and the final examinations as a means of assisting student preparation:

> [A02D202:58,63] *"...they used to have one case, that's how it used to be. So they used to have, during the six weeks, they used to get one case with a consultant, which was a long case.....the decision to change from that system to the two 30-minute cases as it is... was made to more mirror the April and November [Finals] exams, and to say that two cases are better than one..."*

The examination question content of exit-level assessments, to a certain extent, was influenced by and reflected current health issues in the local context. Examiners at Makerere University stressed the need for students to know common clinical conditions seen in their context (popularly termed "Must Know" conditions), especially the ones contributing to maternal and neonatal mortality, and this issue emerged strongly from the interviews.

[B01D210:237-238] *"So we look when we are assessing students, the major contributors of child mortality are one of the examiner generic areas, always examiner generic are nutrition, newborn care, infectious diseases. So… you will find that those areas will always come up mainly for the undergraduate because they need to know and they need to be well versed with the common killers of children, the major contributors.*
So in our exam we try to make sure we will not bring a syndrome or something bizarre for them. We bring the common causes of morbidity and mortality. So we put in a nutrition question, we put in a newborn care, an infectious disease question, child health card for interpretation, something like that. So just to make sure that at least they know the basics of child mortality."

[B02D115:155] *"Maybe one other thing that we look at is the issue of gearing them to know the common things. We wouldn't want them to know about coeliac disease in an exam when they don't know much about malaria. So, somehow even the exam is also a little bit biased towards the common infections because that's what they will actually go out and see."*

[B03Y117:25] *"The way they choose the questions to be set is like, you want to prepare a doctor who is going out, and for us in Obs and Gynae, we try and look at the causes of maternal and neonatal mortality. For example, things like PPH [Postpartum haemorrhage], things like cancer of the cervix, things like birth asphyxia. So, we try to set questions which will target these things. For example you have to make sure there is a question on something like postpartum haemorrhage, you have to set a question on obstructed labour, you have to examine these students so that when they go out there, at least you are sure that probably they are examined on some of these key contributors of maternal or neonatal mortality."*

Similar sentiments about knowing common clinical conditions were expressed by examiners at University of Botswana.

There was also the issue of length of examination time for students: this varied considerably with choice of assessment method. For example, assessing the student in multiple OSCE stations with different examiners was perceived as better than assessing the student on one long clinical case. Seen from a different perspective, it also provided the students more than one single opportunity to demonstrate what they knew and what they could "show how". However, this

[B02D114:107] *"...the one good advantage of the OSCE is the amount of time the student takes to be examined. You know, for the long case, when we did the long case - our generation did the long case - you get there, and you clerk your patient for one hour, then you sit and wait for the examiners, and then they come and stay around for 10 to 15 minutes. You have been in the exam room for about two hours, but you are examined for what, 10 or 15 minutes, whereas for OSCEs, you are in the room for about an hour, or one and a half, and all that time you are being examined. You have an opportunity, you move from one person to another, and you have an opportunity to express yourself."*

Examiner subjectivity or bias was another issue that emerged from the interviews, as a reason for choosing the assessment methods that are currently used. Examiners themselves were aware that this occurred and described in some detail some of the instances that they could recall.

[A03D206:98] *"I've been in discussions where students have failed due to forcefulness of one examiner's personality when they probably didn't deserve to fail. So I think that there are problems, but I think we've got a much better assessment than we used to have, and I think that the knowledge component is already much better because it's much more standardised, it's quality assured, and the people that are marking, there are two people marking each question, which I think has made a huge difference, whereas we used to each share four or five scripts, and then [A03D203] might allow something that I don't, or I feel that they should fail because they did an LP in a child with raised intracranial pressure and she doesn't do that. So, it's inconsistent marking. So I think that the knowledge component is much better, but I think the clinical component can be improved."*

[A09D257:114-115] "*But it might also have something to do with... I'm not sure... I'm just wondering...because as opposed to having one examiner, if you have got two examiners and you have got different cases, one would probably be better able to assess whether the student was just having a bad day with the one case, or they really didn't know anything about the first case but they were safe enough for the second case where they earned a pass in the second case.*
Also just I suppose to have different examiners, because you could have a very strict or an unfair person, you know, and then that

makes up with the second examiner. I mean, that might just be it, and also, to examine someone's skills on just one patient is probably not fair and not good enough."

[B01D107:164-165] *"…in the old system people used to, some people would claim that a particular examiner fails more students than the other one. So when you have got this examiner, you are going to get either low marks or you are going to fail. But now, if all of them are going through the same examiner, I mean there is nothing to say that this one has only failed me or I failed because I went to a hard station or a hard examiner. But all go through the same examiner, so it's more objective.*
The examiner also actually has a guide. They do not ask their own questions. In the old system, the examiner would ask you their own questions, whatever questions they want for that particular case, but now the questions are already predetermined to ask the student, and the student, when he knows it, or has done it, then he passes. But somebody could wake up and ask you what are the genetics of this condition, and you have never heard of them."

[B05Y124:89] *" when they have vivas, they go into the room and here is Dr Z looking at the student. When he looks at the student, the student may start to quake straight away, and now if you add to that there is no structure in the questions, unless you sit down and say We are going to ask this and ask that and ask that, you will find that eventually you are testing the steadiness of the student and the residual knowledge has been driven out by the adrenalin [laughter]. But with the OSCE you know, it's a different thing. You have something which is in place, which is more structured."*

Although they evidently were not aware of the technical terminology, these quotes illustrate that respondents hold a tacit understanding of issues including case specificity and inter-examiner reliability that bedevil assessment on clinical competence.

5.4.3.4 Feasibility as an influence on assessment practices

Another reason for choice of assessment methods related to the amount of time examiners would be involved, as this would detract their available time to provide much needed clinical care for patients.

[B08D135:53] "*One is, just the convenience of it all, in that the long case short case will take a whole day, a whole week to organise, but the OSCE, in terms of its execution is very short. In one morning you have done so many cases. So the examiners, all of a sudden they say, Oh, the exam is half day work, not a one week thing [laughs]."*

In the course of the interviews, examiners did make reference to the issue of resource constraints, but more from the point of view of provision of clinical services and health care rather than in terms of education. Resource constraints with regard to conduct of assessment, however, did not appear to be a feature. On more than one occasion shortages of equipment or medical supplies were mentioned, as the quotations below illustrate. These are included here to emphasize the challenging conditions that health care providers in many parts of Sub-Saharan Africa still work under.

[B01D210:207-209] *"Just to add to that, of course that impacts on our training, because what happens is that if the things are not there, then the students cannot engage, even when they are being trained. You cannot engage, because if I give you an example, you know we have many universities here, but they come here for internship. So one of the students, we asked them, Oh, this child has been transfused many times, so we are talking to the intern and saying, Surely this time why don't you do a blood film before you give blood? Just as you do a blood slide, just take off a film, and they are, like, What is that? So we are asking him, you mean you don't know what a blood film is? Like, I can't do a thick film, it's not my role, and I have never done one. So we are like, you mean you don't do practicals in your setting? They say no, they stopped them because there was no person to teach them in their university, the lab person, but there were also no resources for that. So that gap impacts on our training, because they will go out of the system without having known what the ideal is, because they have never used it. So they will go with what they know best."*

[B02D114:257] "*…on the issue of policies and funding, one time we finished making the guidelines, we produced guidelines and we said okay, very good. So I met the officer afterwards - I was part of the writing team - she said, "When are we launching - when are we getting copies to distribute to the whole country?" So, give me money [chuckles], but she didn't have money. The guidelines were there in electronic form, they put on a website, but how many health workers can go to the website? They produced about ten books to start with [General laughter] for the people who participated in the writing. So, I had a copy, but I knew everything in the book [laughs]. Now they have been disseminated, but you see, it took a while by the time that came out…"*

[B03Y117:118] *"Supplies… things run out, like, gloves, catheters, medications. Sometimes if you need to give treatment, you have to give treatment which is available. Instead of your choice, you give what is available. I want Ceftriaxone, they say, Oh, we have Ampicillin. Okay, give Ampicillin, but the resistance to it is high, so this patient may not improve. You know what to do, but you don't have what to use…."*

[B03Y117:122] *"We have magnesium sulphate, but sometimes, there are occasions when it is out of stock. When you don't have magnesium sulphate, you don't have labetalol, you don't have hydralazine. Sometimes, some rare occasions, these few occasions that they can be out of stock where you ask the patients to go and buy...."*

[B03Y117:134] "*...Power supply, very minimum, because the hospital has a generator. It has a standby generator. If it goes off, the hydroelectric power, then the standby generator goes on. But, so, things like equipment. Right now, the operating tables we have in the theatre are faulty. They are obsolete, according to the manufacturers [chuckles]. They are obsolete, according to the manufacturers! We were just surfing them on the Internet, the manufacturers said last month, they are 30 years ago, and those parts are no longer there. So, forget about that type of table [chuckles] in that theatre..."*

5.4.4 Expectations for graduates

The question of what clinical competence was considered appropriate for medical graduates from these medical schools in relation to pertinent health-related issues was also explored during the interviews with examiners.

5.4.4.1 Clinical competence

There was general agreement among examiners from all three study sites regarding their understanding of clinical competence and what expectations they had of their students. The common points that emerged were that clinical competence involved mastering the steps required in the complex process of clinical management. This included the skills of history taking and physical examination, integrating the findings in order to make a diagnosis, followed by initiating appropriate investigations and treatment. A good knowledge base was required in order to perform these tasks, and students also needed to become proficient at basic procedural skills and in communicating with other health-care providers, patients and their families.

[B01D210:36] *"...we have actually also been assessing competence, we have tried to think of also attitude involved, other than knowledge and their skills. Attitude has also been one of the competencies, and of course we assess it in the form of their professionalism, ethical, are they doing all this together with the right ethics."*

A number of examiners placed emphasis on aspects of professionalism, having the appropriate attitude and ethical behaviour, and lifelong learning skills were also

deemed to be part of clinical competence. However, while it was evidently expected of students, it was for the most part not clear how the assessment of these components was carried out.

[B04Y120:37] *"Yes, it is very subjective, and you find that the summative, this final assessment may not be very good. So much of it is assessed in formative stages as students go on the department. Like, you call them for teaching, whether they come, when they respond, how they communicate, how they interact with each other during the training sessions, or during the tutorials, or in any other areas where they interact as forms of teaching. Much of it is by - actually, the training - is by role modelling. So the assessment is really, we try, but I don't know whether we actually… assess professionalism."*

[C02D261:202] *"…I think that's why we have also modified our logbooks as well. We didn't used to mark them on professionalism, we didn't used to mark them on attendance, we didn't used to mark them on certain things, so now we actually have it in there because they were ducking out, and because it's not important, but it actually is important. So again, it's trying to stress the importance."*

[C03D162:36] *"Professionalism is a core component of clinical practice. So, professionalism for me looks at how one relates to others, how dependable one is, how accountable one is. I am not just accountable to my colleagues, I'm not just accountable to my patients, I am also accountable to the guardians of the patients. I am also accountable to the whole system that has brought me here. People look at me here as a paediatrician. There is a system, and I must be accountable in that system. I must be accountable in what I have done, in what I have said, you know. I must be accountable in all those things."*

Examiners expected their students, by the time of graduation, to have mastered the basic steps in the clinical management of patients, and also to recognise when and how to ask for assistance. A key feature of clinical competence that examiners looked for in their graduating doctors was whether they could be deemed to be "safe to practise". Examiners wanted to be assured that the patient care provided by the graduating doctors would be appropriate and adequate, with patients in safe hands.

[A03D104:80] *"…at the end of the day, I think the final decision or question we're trying to answer for ourselves in terms of whether we're going to pass a student or not is whether we think they're safe to practise."*

[A08D252:48] *"…at the end of the day, after the exam, you must be able to send them out and they are safe. That's the question when I answer [sic] – are they safe?"*

[B03Y117:049] "*Clinical competence involves someone being able to make a diagnosis. Someone being able to make a diagnosis and someone being able to know that I can manage this, and this one has to be handled by a different person.*"

However, concern was also expressed that students, especially with reference to weaker or borderline students, could pass although perhaps they still had gaps in terms of their competence. This seemed to be contradictory and highlighted an issue that examiners appear to find challenging to deal with:

[A03D104:80] "*So if we think they are safe and they won't kill a patient, even if their knowledge is shocking, we tend to sort of pass them……I don't think that if we look at the assessment, and obviously from an institutional point of view, in society, like we only want to pass people that are competent. As an assessment in terms of grading performance of students, so actually differentiating well between above average and below average students in an objective fashion, I don't think it particularly delineates the middle group particularly specifically. But I think the discussions often around a pass fail candidate become very subjective and reduced to are they safe or not, without much…. I don't know…. and I'm not sure what we base that on.*"

[A03D203:83] "*…sometimes in the written paper you have to be careful because the student can actually pass the question but do dangerous things. I think we should build in some or other rule there that if they answer it in a dangerous fashion they get zero for the question.*"

Concern was also expressed about what pass mark was deemed acceptable in terms of the level of clinical competence:

[A08D252:41] *"So, clinical competence, obviously for us it means that they have to pass a written exam as well as pass a clinical exam, but it's hard to think that they only need to get 50% to pass that exam, because what are we saying? That they only know 50% of the work, and then they are competent doctors, and sometimes for me that doesn't sound good. You actually want them to get a better mark than just 50%, so how do we decide on that pass mark? That is a problem for me. But for me, okay, both clinical and written exam should be up to standard."*

It was notable that none of the institutions participating in the study used standard-setting practices.

Some examiners were aware of gaps in the exit-level assessments and this perhaps hinted at what they felt should be included as part of assessment.

[A03D206:96] "*I mean, the downside is we never test history, we don't test communication, we don't test all those graduate attributes at all. We don't ever assess that, and I think those are as important in graduating a doctor. Now, I don't claim to know how one is going to assess that. I know people are building it into assessments. But that is a problem, and I think that it would be really nice to be able to do an assessment, for example, in an outpatient type setting where you have children that come in with acute illnesses. I mean, you are looking at Millennium Developmental Goals. I don't think that any of us ever align our assessment with Millennium Developmental Goal aims or anything like that."*

[A08D253:60] *"Unfortunately the only skill that we actually examine them on is actually the examination skill. The skill that goes further than examination skill, like interventional skills, we do not test on. I think the university has the skills lab that has those tests they do in the skills lab, in the earlier years, that actually tests them on like basic sort of interventional skills. They test them on that, but we don't test them on those type of technical things."*

5.4.4.2 Acquisition of skills

By the time of graduation, certain basic skills were expected of students from each discipline. For example, in Obstetrics & Gynaecology, this included the ability to monitor labour and to conduct a normal delivery, to perform and suture an episiotomy, and to conduct a gynaecological examination. It was essential to be able to manage obstetric emergency conditions which contributed to common causes of maternal deaths, such as postpartum haemorrhage, pregnancy-induced hypertension and eclampsia. In Paediatrics, emergency care of the newborn (which included resuscitation skills) was important. Students also needed to be able to pick up abnormalities at birth and identify common neonatal problems.

[A05D247:042] *"…so what we teach them is to pick up once a baby is born if there is anything abnormal, anything acute, serious. They need to be able to identify when a baby is ill and when they need to be referred and how to stabilise the baby."*

Examiners believed these skills were acquired through a process of initial observation, learning to perform them, and then followed by practising. With more complex surgical procedures handled by the more senior members of the clinical team, such as caesarean sections and laparotomies (in Obstetrics & Gynaecology), final year students and interns were expected to assist in the operating theatre in order to familiarise themselves with the indications for these procedures and what to do.

However, this examiner's remarks seem to suggest that much of the assessment focusses on "Knows" and "Knows How" tiers of Miller's 'pyramid' with less emphasis on "Shows How" or even "Does", and an awareness that there is a gap in the assessment:

[B02D114:153] "*The written exam is structured to assess knowledge, assess aspects of public health, assess management, assess also a lot about how they can go through that, put all their ideas together. That's why there are multiple choice questions which test recall, then clinical oriented questions which test the practicality of their knowledge, managing clinical cases. Then there is also usually one or two cases on public health aspects. It could be immunisation, it could be something on child health, it could be adolescent health that they are able to bring out all these competencies."*

As indicated earlier in Tables 5.7 and 5.8, it was evident that the exit-level assessments at all three study sites did not have any examination questions in the "Does" tier of Miller's 'pyramid', yet examiners in interviews mentioned procedures and other skills that students were expected to learn and practise in the ward and clinic settings:

[A06Y149:071] *"We would certainly require a final year medical student to be able to manage acute haemorrhage and a ruptured placenta and hypertensive disorders as an absolute minimum. It is 'Must Know' knowledge."*

[A07Y251:458] *"Even a basic vaginal, if someone is in labour, you should be able to do a basic."*

[B03Y117:057] *"…they should know that this is how I resuscitate a patient, this is how I work up a patient who is going for such an operation, and they should know they have to take bloods for group and cross match, put up a line. Of course, all those skills are polished during internship, but they should know that this is supposed to be done. It is once when someone goes out and they cannot make a diagnosis of an ectopic pregnancy, they cannot know that this is obstructed labour, we have to catheterise, obstructed labour, we have to give antibiotics. So, they have to know what to do…"*

Examiners at University of Botswana had very clear expectations of the skills that their students needed to acquire, especially with regard to skills that would be pertinent to their future practice and provision of health-care:

[C02D261:066] *"…at the fifth year level we expect them to do some basic procedures, but I think some of them go beyond that, and the very keen ones are*

> *functioning as a medical officer or maybe even a junior MMed resident level."*

[C04Y263:53-54] *"At year five we expect them to be able to conduct a normal delivery, to monitor labour and conduct a normal delivery. Also they should be able to assist and later perform an episiotomy and repair it, be able to stitch lacerations, manage a bleeding patient. They are required to participate in managing obstetric emergencies. So in this rotation they clerk and present cases of the obstetric matters they managed. I am talking of things like our main problems, in particular the post-partum haemorrhage, eclampsia etc. …*

Then regarding, abortion-related haemorrhage, they are required to perform at least ten evacuations. So they are able to quickly empty the uterus, control bleeding and offer comprehensive post abortion care. We also require them to assist in theatre; they have to assist at least five caesarean sections, so that they are familiar with what the indications are and what to do. They also assist at elective gynae operations, at least five. This includes laparotomies for ectopic pregnancy, cancers, fibroids. So they have adequate exposure."

As discussed earlier in this chapter (in Section 5.3.1.1 and with reference to Tables 5.8 and 5.9), the expectations University of Botswana examiners had about the skills their students needed to acquire was reflected in the medical- and surgical-based OSCEs at exit-level. Deductive analysis showed all the OSCE stations were at the "Shows How" tier of Miller's 'pyramid'. The use of a logbook to record procedures that were performed (such as normal vaginal deliveries, performing and repairing episiotomies, performing evacuation and curettage of spontaneous abortions, as well as assisting in other specified obstetric and gynaecological surgical procedures) was also noted in Section 5.2.1.2. This suggests that logbooks, as used in this instance at the University of Botswana serve an important function in the assessment of these skills which would be a crucial contribution towards making decisions on the clinical competence of these students.

5.4.4.3 Preparedness for internship

Immediately following graduation, intern doctors are required, by law in their respective countries, to work under supervision for a minimum of one year in Botswana and Uganda, and two years in South Africa before they can be fully licensed medical practitioners for independent practice. Examiners recognised that newly-graduated

interns would be more proficient in theoretical knowledge and skills but less experienced in practical ("hands on") aspects of clinical management.

[B03Y117:57] *"…So they have to know what to do, and in internship, they can practise."*

[A06Y148:016] *"At the end of the final year as an undergraduate medical student, the students that qualify, we would like them to have the competencies to be able to perform the tasks that they will have to do when they are interns…"*

The internship was seen as a period of preparation for future practice, and interns were required to take on more responsibilities in patient care. The period of internship would therefore provide an opportunity for their clinical competence to be developed and their skills enhanced under the careful guidance of more experienced colleagues and senior consultants.

Also as part of preparation for internship, students in their final year of training in medical school would already have been expected to shadow the hospital interns in order to start learning the role and what is required in order to fulfil that role.

[C04Y263:55] *"…year five [Final year at this institution] students shadow as intern doctors on the ward. So shadowing as interns will mean that they do the pre-round, do vitals, evaluate the patient and draw a management drug plan. Thereafter we will see and discuss the patients with them on the ward round as they present. They are able to draw bloods for basic tests, order a scan which I counter-sign. So we make them try to do key internship roles, so that at the end of the day, they are on a fair footing and well equipped for their tasks."*

5.5 Concluding remarks

This chapter has detailed the findings emanating from the study. The methods and content of assessment in two disciplines at three medical schools in Sub-Saharan Africa have been explored. Both the needs of the region and appropriateness to exit-level assessment have been used to interrogate the findings. In the following Chapter 6, the next steps in synthesizing these findings will be through interpretation and discussion, ending in the significance of the study and recommendations for future practice and research.

Chapter 6

Interpretation & Synthesis

6.1 Introduction

The need for more research in education to guide teaching principles as well as assessment has been acknowledged for quite some time now (Gage, 1989; Davies, 1999; Harden et al., 1999; van der Vleuten et al., 2000; Prideaux and Bligh, 2002; Albert, 2004; Cook et al., 2008). Research provides the evidence to inform future education policies and practices, and also contributes towards ensuring standards are maintained and that there is accountability. The relevance, applicability and quality of the educational research are also important (Davies, 1999; Eva and Lingard, 2008).

This study sought to explore exit-level assessment and the extent to which assessment practices focus on clinical competence. More specifically, the study examined the validity of the exit-level assessment practices in determining the clinical competence of medical graduates for selected health-related issues at three medical school in Sub-Saharan Africa, a region of the world characterised by a high burden of disease and a shortage of doctors as well as other health-care professionals. A brief recap of the background to this study is outlined here.

The training programme for doctors in most medical schools is a long one, with the curriculum spread between four to six years. Assessment occurs at various stages during the training programme to ensure appropriate levels of achievement have been attained, and at the end of the training programme, to certify competence and fitness to practise as doctors (Epstein, 2007). In Chapter 2, attention was drawn to the fact that assessing competence is a complex and multidimensional undertaking (Miller, 1990; Carraccio et al., 2002; Frank et al., 2010a). Exit-level assessment is important because medical schools have a responsibility to ensure that graduates have attained a level of competence (Epstein and Hundert, 2002; Shumway and Harden, 2003) for reasons of public accountability and patient safety. Exit-level assessments need to be suitably robust in order to differentiate between the sufficiently and insufficiently competent

individuals (General Medical Council, 2009; General Medical Council, 2011; Liaison Committee on Medical Education, 2013). Therefore, criteria for sound assessment are needed to guide the process of exit-level assessment.

As indicated in Chapter 2 (Section 2.1.1), assessments, particularly at exit-level, need to be aligned with intended outcomes of curricula, which in turn, need to be aligned with the health-care needs of the country. This reciprocal relationship between educational objectives and health-care needs is an important one. In highlighting the divide between the education of health professionals and the health-care systems of countries, Frenk et al. (2010) have drawn attention to the necessity for changes in education of the health professionals in order to provide graduates who are better equipped to deal with the challenges of health-care in the 21st century. This imperative is even more relevant in resource-poor contexts such as Sub-Saharan Africa where meagre resources impede the provision of appropriate and adequate health-care.

The Sub-Saharan African region has consistently suffered from a crippling lack of resources such as human capacity and an inversely high burden of disease in comparison to other regions in the world (World Health Organisation, 2006; World Health Organisation, 2014). Shortages in infrastructure and equipment in addition to human capacity and other resources required by medical schools to provide training for health-care professionals were highlighted in Chapter 3. With demands being placed both by the provision of health-care and the training of doctors and other healthcare professionals, the allocation of strained resources becomes a precarious balancing act where the provision of health-care services is often inevitably foregrounded. This can lead to particular challenges faced by faculty who are also educators as they may not be able to focus as much attention on education as they may wish to do.

Education for academic faculty in medical schools in Sub-Saharan Africa, as in other institutions worldwide, is one of the multiple roles they need to fulfil, which roles also encompass research, and clinical practice and service. These roles are typically in tension with each other (See Section 3.4). Any consideration given to education may tend to focus more on the immediate aspects of teaching and delivery of the curriculum rather than on aspects of assessment. This has implications for the assessment of clinical competence of doctors in Sub-Saharan Africa and potentially impacts on the quality and appropriate health-care and patient safety.

The remaining sections of this chapter will provide a synthesis of the findings that have emerged from this study, then leading to implications for practice and lessons learnt. Some issues emerged that have the potential for future research study and recommendations have been made for their consideration. These include exploring reasons for the continuing use of certain assessment methods (such as essays), investigating the feasibility of adopting workplace-based assessments, further exploring what changes need to be made in Sub-Saharan African medical schools in order to shift exit-level assessment from the lower to higher tiers of Miller's 'pyramid'. The chapter will round off with a discussion of some of the limitations of the study as well as some personal reflections. But first, some reflections about the two analytical tools used in this study and which featured prominently in Chapter 5, are provided.

6.1.1 Reflections on Analytical Tools Used

Assessment remains a core component of the teaching and learning project and as such requires ongoing attention. Generating data by means of document analysis, interviews and observation provided an array of findings that required careful analysis. The use of appropriate analytical tools was critical and those used in this study were described initially in Section 4.6. Miller's 'pyramid' was used in relation to assessment practices and to the appropriateness of the assessment content for exit-level (See Section 5.3.1 for the results), and Disability Adjusted Life Years (DALYs) were used in relation to the content validity of exit-level assessment questions (See Section 5.3.2). These tools have not been used in this context before and some reflection on their value in the study is required.

6.1.1.1 Miller's 'pyramid'

Using Miller's 'pyramid' as an analytical tool for this study was a useful exercise providing critical insights as to the level at which the different assessment events were pitched. However, its application was not a straightforward exercise. As analysis progressed, it became apparent that categorisation of examination questions required a combination of educational knowledge and clinical experience to decide to which tier of the pyramid a question should be assigned. These judgements or interpretations of the questions were inevitably subjective, and this needs to be taken into consideration when interpreting the outcomes of this study. Nevertheless, every attempt was made to

meticulously document the process of analysis which took time and was labour-intensive. The use of Miller's 'pyramid' as an analytical tool further revealed that its application to individual examination questions was more accurate than its application as an overall global rating to any single assessment method. The analysis highlighted how varied the assessment tasks can be that are devised within the framework of an assessment method like an OSCE. It was apparent that applying Miller's 'pyramid' broadly to assessment methods like "OSCE" and "MCQ" could misrepresent – indeed, overestimate – the level at which students are being assessed. As is evident in Tables 5.8 and 5.9, not all OSCE questions were situated at the "Shows How" tier of the 'pyramid'. Rather, a more fine-grained analysis is required to ascertain the level of assessment. This has implications for the design of assessment systems in medical schools. It also has implications for regulators, who cannot merely look at the "big picture" of what assessment is in place but must necessarily drill down to the level of individual assessment tasks to truly understand assessment in a school and whether that assessment provides the information required for licensing and to ensure public safety.

6.1.1.2 Disability Adjusted Life Years (DALYs)

While Miller's pyramid provided an analytical tool for reviewing the appropriateness of the nature of the assessment method, it was necessary to find a suitable frame to review the content of what was being assessed. In this study, therefore, DALYs were used as a second analytical tool and, as was the case with Miller's, this also constituted a novel approach. The choice of DALYs as an analytical tool is open to question. Even though DALYs have proven to have greater utility than Millennium Development Goals (MDGs) for this purpose (Section 4.6.2), it is by no means a perfect match. Nevertheless, in the absence of alternative tools for analysis of assessment content, it proved to be a useful option as can be seen from the discussion in Chapter 5 (Section 5.3.2).

Ultimately, however, it highlights the challenges faced in trying to provide a suitable mechanism to demonstrate alignment of the curriculum to the health-care needs of the local community. It also draws attention to the need for relevant contextual burden of disease data to inform curriculum development in developing countries. The same burden of disease and resource limitations that impact on healthcare delivery and education also impact on the health system and its capacity to collect and analyse

relevant data. Currently available metrics are simply not adequate to the task of informing curriculum development or review. This means that these processes are subject to the vagaries of the perceptions of the (typically tertiary-centre based) care-givers themselves about what is important and relevant, which can lead to curriculum overload and inappropriate preparation of students for the demands they will actually face. The ramifications of this extend beyond the confines of the curriculum. In addition to involving curriculum designers and those involved in curriculum planning, it could also have an impact on what (local or external) funders should resource.

6.2 Synthesis of findings

Notwithstanding the caveats described above, this study provides a crucial and unique exploration of assessment – its practice and its relevance – within a particular context, namely, that of under-resourced health-care contexts. A number of key issues emerged from the study that have relevance for the way in which assessment is both conceptualised and implemented, and will each be discussed in turn. These issues relate to assessment in practice, recognise examiners as key players, and explicate the influence of context from an institutional perspective on assessment practice.

6.2.1 Assessment in practice

Several points are considered in this subsection on assessment in practice. These include the assessment methods used, the integrated approach to assessment that was used at one of the institutions in the study, the assessment content in relation to local health-care issues, and the weighting of marks.

6.2.1.1 Assessment methods

Although institutions used a selection of methods for exit-level assessment, the majority of what was being assessed by the assessment methods being applied were found to be at the lower tiers of Miller's 'pyramid' (See Tables 5.8 and 5.9) with none at the "Does" tier. It could be argued, therefore, that students were, not being given sufficient opportunity to demonstrate the extent of their knowledge and skills, and that current assessment practices do not provide for relevant data to be gathered and decisions to be made about competence. Exit-level assessment should allow for decisions about clinical competence rather than just knowledge – the latter being the 'type' of evidence that

would be generated at Miller's 'lower' tiers. In the context of this study it could be argued that the evidence gathered through assessments was insufficient to allow accurate judgements to be made about the clinical competence of the students and for conclusions to be drawn about their ability to practise medicine and in particular to handle clinical situations arising from the context in which they will be practising on graduation. This is an important consideration. In resource constrained settings, if the health workforce is poorly prepared and therefore unable to cope with the health-care needs of the community, the quality of health-care provided would be compromised. Resource limitations result in understaffing of healthcare facilities which in turn means that on-the-job training is limited. The layers of supervision and training available in other settings are simply not possible when trying to deal with the burden of disease and a limited workforce working in poorly resourced facilities. Newly minted doctors need to be able to "do" not just "know".

It is now largely accepted that there is no single assessment method that comprehensively assesses clinical competence by nature of the complexity of competence (Miller, 1990; van der Vleuten and Schuwirth, 2005; van der Vleuten et al., 2010). As discussed in Section 2.4.2 exit-level assessment practices reported in the literature indicate that ideally a selection of methods is widely used, as the use of multiple methods provides a better overall picture of the performance and competence of the student. The exit-level assessment methods in use in the institutions that are the subject of this study were found to be generally in keeping with worldwide practices, in terms of using a selection of methods and in choice of methods. However, the methods of assessment such as essays, long cases and certain forms of oral that have become less used elsewhere (Fowell and Bligh, 1998; Davis and Karunathilake, 2005) were also found to still be in use at the institutions included in this study. This could be problematic in light of their documented limitations (See Section 2.5 where various summative assessment methods were discussed). These methods, therefore, warrant further comment here with regard to their application in the Sub-Saharan African context. It could be argued that for the time and resources needed, there are more effective and efficient means of gathering data to make decision at exit-level.

For example, **Essays** formed an integral part of the written component in the final examinations for Obstetrics & Gynaecology at Makerere University. In this high-stakes

assessment, two essay questions (one each covering an obstetric- and a gynaecological-based topic) contributed 16.7% proportion to the mark in that examination and to a tenth of the overall final mark in that discipline. Furthermore, analysis of the level of questioning and content revealed that they only tested to the "Knows How" tier of Miller's 'pyramid'. The way that essays are used in this context therefore seems to be characterised by the very flaw that are highlighted in the literature advocating against the use of essays. Essays are easy to set, they require more time to answer, are difficult to mark, and they have low reliability (Fowell and Bligh, 1998; Schuwirth and van der Vleuten, 2004). Hift (2014), drawing on an extensive literature review, concluded that the purported belief that essays have greater validity than multiple choice questions is actually not supported in the literature. The way that essays are used in this context lends credence to Hift's argument. If gathering evidence of graduates' competence is a goal of exit-level assessment, then a strong argument could be put forward that the considerable time that marking essays takes could be better utilised in a resource constrained setting and focussing on a more practice-based approach. Future research will need to explore and understand why such methods are still used at exit-level, this to help inform efforts to improve exit-level assessment.

Orals are an assessment method with a long history of use, although they have been criticised for having low reliability and validity, being time consuming and have been deemed to add "little of value" to the assessment process (Norman, 2000:208). Yet they continue to be in use because "breaking with tradition" is an enormous challenge (Davis and Karunathilake, 2005:294). Although orals are not officially listed as an exit-level assessment method used in this study, it was evident from observations made in the Stellenbosch University Obstetrics & Gynaecology OSCE/OSPE that the student-examiner interactions at the OSPE stations are essentially a semi-structured oral examination (Section 5.2.1.2), rather than a conventional OSCE as described in Section 2.5. This anomaly was subsequently confirmed during the interview with the clinical clerkship coordinator. Orals usually take the form of a face-to-face encounter between the examiner(s) and student, in a setting such as an examination hall, away from patients (Davis and Karunathilake, 2005), as was the case at Stellenbosch University. The four "OSPE" station questions, constituting 16.7% of the final mark (See Table 5.8), focus on the clinical management of obstetrics or gynaecological conditions or situations. However, as they are more of a theoretical nature only requiring verbal

responses from the students, it can be argued that they test only to the level of "Knows How" in Miller's 'pyramid'. This could, therefore, have an impact on the validity of decisions being made about clinical competence at exit-level.

The use of orals or similar approaches in the face of evidence that would counter their use raises questions as to why they are still in play. In several study interviews, examiners could not always explain why certain assessment methods were being used, but indicated that practices that had been in place for some time were simply accepted and continued (See Section 5.4.3.2). Maintaining traditions in assessment practices that have been inherited from senior faculty without question suggests there is an institutional culture at play, and this will be discussed further in Section 6.2.3. The fact that many health professions educators mirror their practice as teachers on their own experience of being taught has relevance in this context. Some of the more senior examiners had themselves experienced the orals format during the course of their own training (See Section 5.4.3.2). It is plausible that this could have been a contributing factor or influence to its retention as an assessment method as examiners drew upon their own experiences and use of methods they were familiar with, although this was not explored in-depth as part of this study.

So, are essays and oral appropriate as choices of assessment methods to assess the competencies needed for a doctor in resource-constrained settings in the 21st century? It could be argued that the lack of resources in these settings has had the further invidious impact of hampering the acquisition of up-to-date knowledge about assessment by busy clinicians. The resource constraints perhaps allow the power of 'tradition' to play out. So there is a matrix of influences that ultimately impact on the judgements that are being made about competence, and this needs to be borne in mind when considering the extent to which exit-level assessment methods and practices are valid in determining the clinical competence of medical graduates and how to improve this.

Methods from the "Other" assessment category (See Section 5.2.1.3) used in clinical clerkship assessments but not in the final examinations could be regarded as providing better measures of the skills and attitude components than some of the more traditional approaches. "**Other**" assessment methods include those that do not fit into either "written" or "performance-based" categories (See Tables 5.2 and 5.3). They are based

on clerkship activities, such as the seminar presentations and Problem-Based Learning (PBL) sessions at the University of Botswana. The PBL sessions constituted 10% of the final mark, and seminar presentations (which feature in Obstetrics & Gynaecology) constituted a further 10% (See Tables 5.2 and 5.3). According to the marking schedule in the student handbook, 5% of the marks awarded are based on their knowledge of the topic being discussed, and the remaining 5% on "participation" (this includes punctuality, and peer interaction in the PBL group). These activities provide rich opportunities for students to develop and foster skills that would not necessarily be developed (or assessed) in other more conventional approaches, such as learning and practising communication skills and also in the giving and receiving of feedback, and developing appropriate professional behaviour.

All three study sites also make use of some measure of attitude or professionalism, using ward evaluations, ward marks, professional conduct and continuous assessment. These generic or graduate attributes feature increasingly in medical curriculum discussions (Frank and Danoff, 2007). Although not clearly defined in the study documents that were analysed, they appeared to encompass observations of behaviour but were not necessarily explained nor was their significance described. Typically, these "other" assessments contributed to a smaller percentage of the final marks (See Tables 5.2 and 5.3), suggesting there was less importance or emphasis attached to them in comparison to summative assessment of knowledge and cognitive skills as well as psychomotor skills. However, this does not appear to match the comments made by examiners and clinical clerkship coordinators in the interviews in which they expressed a strong belief in the value of lifelong learning skills, professionalism and having the appropriate attitude and ethical behaviour as part of their expectations of clinical competence. If professionalism were to be regarded as equally important and relevant irrespective of the setting in which medicine is being practised, then it needs to be afforded the emphasis it deserves and this also needs to be reflected in the final examinations. Exploring the existing literature on professionalism would help guide its practice of assessment (what to assess and how to assess) and points to an area for future enquiry (Ginsburg et al., 2000; Steinert et al., 2005; Veloski et al., 2005; Cruess et al., 2006; Steinert et al., 2006; Jha et al., 2006; Rees and Knight, 2007; van Mook et al., 2009; Wilkinson et al., 2009; O'Sullivan et al., 2012).

The relatively low weightings for aspects of professionalism and attitude seen at all three study sites are probably in keeping with the trend seen in other medical curricula (Howe et al., 2004; McCrorie and Boursicot, 2009). Given the health-care demands of the Sub-Saharan African context that have already been deliberated in Sections 3.1 and 3.2, dealing with the huge burden of disease in this region would still be a priority and it would seem reasonably justified for more emphasis to be placed at exit-level on testing cognitive abilities and psychomotor skills that graduates need to be equipped with in order to function as clinically competent doctors in this context. However, in such under-resourced, pressurised environments, having a positive attitude and a professional stance are equally crucial and need to be cultivated and encouraged.

6.2.1.2 Integrated assessment

In contrast to the conventional discipline-based approach adopted at the other two study sites, the University of Botswana used **integrated assessment** (see Section 5.2.1.2). Although this refers more to an approach or model to assessment rather than methods per se, it warrants some discussion here as it still has a link to the methods used. To start with, a clearer definition of "integrated curriculum" would need to be in place first, such as the one proposed by Brauer and Ferguson: "a fully synchronous, trans-disciplinary delivery of information between the foundational sciences and the applied sciences throughout all years of a medical school curriculum" (Brauer and Ferguson, 2015:318). It was not discussed earlier in Chapter 2 as a standard or regular feature of assessments although it appears to be gaining ground as a model (Brauer and Ferguson, 2015). The use of the term 'integrated' was not clearly defined in the available institutional documents but appears to refer to a single assessment event that includes questions that focus on different clinical disciplines and are thus not specific to a single disciplinary focus. Observations of the examinations conducted indicate that different clinical disciplines all contribute questions to the written assessment papers and the medical-based and surgical-based OSCEs which constitute the performance-based assessment component. The corresponding model answers for the questions are provided and marked by examiners from the respective individual disciplines. In other words, the input from various clinical disciplines is combined to make a composite

examination, but essentially, the assessment of the student still takes place at individual discipline level.

The integrated assessment currently taking place at the University of Botswana appears to pose a fundamentally different challenge to exit-level assessment in comparison to the assessment taking place at Stellenbosch and Makerere Universities. Referring back to Tables 5.8 and 5.9, and focussing on the OSCE in particular, the number of OSCE questions in the final assessment at the University of Botswana was smaller compared to the numbers at the other two institutions; these questions carried a proportionately larger percentage of the final marks and, in addition, all the questions featured on the "Shows How" tier of Miller's 'pyramid' whereas a number of OSCE questions from Stellenbosch and Makerere Universities were to be found in the lower tiers. This could be as a result of the integrated assessment, but could also be because examiners at the University of Botswana were designing more appropriate assessments.

A pertinent question at this juncture is whether integrated assessment would be a better way to determine whether students in Sub-Saharan Africa are competent at the end of their medical training. Perhaps this could be considered in light of the setting in which the graduates will be working, as another aspect of integrated assessment relates to the workplace context. Doctors need to be able to deal with undifferentiated clinical presentations. The presenting history of the patient is not always clearly defined and requires the doctor to gather further information (starting with history taking and physical examination, and possibly conduct laboratory and medical investigations) in order to arrive at a diagnosis and to formulate a management plan. This complex process of clinical management was recognised by examiners in the interviews when discussing their understanding of clinical competence (see Section 5.4.4.1). Using the example of a case presenting with abdominal pain as an undifferentiated problem, and a team of examiners from different clinical disciplines, an integrated assessment would involve assessing the student's overall approach to this case from different perspectives, and this could be utilised in the long clinical cases. Alternatively, OSCEs could equally be modified to provide a more integrated examination involving input from various disciplines, "to link theory with practice in order to reflect the reality of professional practice" as described by Hudson and Tonkin (2004). The use of undifferentiated clinical presentations may therefore offer a more authentic impression and more closely

represent what actually occurs in the clinical setting. This may well have been the aim behind University of Botswana's efforts to have 'integrated' assessment at exit-level. It was evident from the interviews (See Section 5.4.1) that during the design process, faculty came to the decision that they had to design an integrated examination. As University of Botswana is a much younger medical school than Makerere and Stellenbosch Universities, there was no existing practice to constrain the design of the assessment system. It is also more likely that in only recently designing a medical curriculum, the designers were more likely exposed to and adopt practices that reflect a contemporary understanding of best practice.

The use of integrated assessment appears worthy of further investigation for use in a resource-constrained context such as Sub-Saharan Africa, and these implications will be discussed further in Section 6.3.

6.2.1.3 Assessment content

This discussion on assessment methods thus far has focussed on the range of methods being used, their appropriateness and whether these have enabled students to demonstrate their clinical competence adequately. Throughout this study it has been clear that the content being assessed is as important as the method in the effort to ensure that faculty are able to make meaningful decisions relating to the clinical competence of graduates for practice. It was therefore crucial to find a mechanism that would enable the scrutiny of content. A key element of this study was that the content being assessed (and therefore, it could be assumed, being taught), would be relevant to local context and prevailing burden of disease.

Assessment content in the final examinations was matched with the burden of disease as described by region-specific causes of Disability Adjusted Life Years (DALYs). Assessment questions were generally found to be appropriately focussing on relevant contextual health-related issues (Table 5.16). However, a number of examination question topics obtained from the two clinical disciplines at all three institutions studied that related to relevant health issues in their respective settings also did not feature on the DALYs tables (See Table 5.17). The question arises whether the proportion of examination question topics relating to DALYs should be greater than they are, since DALYs provide a measure of the contextual disease burden and it would therefore seem logical to focus on them. However, doing so may be at the expense of

other equally important topics which do not feature on DALY tables (which will be discussed next). Consideration also has to be given to the challenge of ensuring comprehensive coverage of relevant topics from the vast field of clinical medicine for the limited available number of examination questions, and in this regard, blueprinting may help (See Section 6.3 later).

As pointed out earlier, there was also assessment content in the final examinations that did not have any relation to leading causes of DALYs (See Table 5.17). These question topics relate mainly to issues of "normal" health-care rather than disease conditions, which are critical given that one does not want to adopt a purely curative approach, and are in inverse proportion to the examination question topics that feature in the DALYs tables. At this point, therefore, a brief digression to consider the basic content of a medical curriculum is required to assist in understanding the relevance of the inclusion or omission of question topics from the DALY tables.

Curriculum relates to the course of study or syllabus and comprises content, teaching and learning strategies, assessment processes and evaluation processes (Prideaux, 2003:268). The content of a medical curriculum would include a foundation based on the study of normal structures, function and behaviour of the human body, how these are affected by disease conditions leading to the study of abnormal structures, function and behaviour. In building upon these foundation blocks, students would learn about the diagnosis, treatment and rehabilitation of diseases processes that comprise the curative approach to medicine. Alongside this approach, aspects of health promotion, disease prevention and the delivery of primary or essential health-care, which are equally important in the practice of medicine, would also need to be learned. A logical next step would be ensuring that any assessment that takes place is aligned with the curriculum (See Section 2.1.1). Assessment should therefore entail testing students on the "normal" healthy state as well as the "abnormal" disease conditions, and on aspects of curative as well as preventive health-care

Therefore now, with regard to the choice of DALYs as an analytical tool for the content of examination questions, it is not surprising to note that DALYs have been devised with a focus on disease conditions. It has provided an excellent approach to the disease-based curriculum, but has also highlighted its limitations for use in the remaining non-disease-based curriculum.

It should further be noted that some question topics relating to specific diseases or conditions did not feature high on the causes of DALYs lists. Examples include gynaecological malignancies (such as cancers of the cervix, ovary and uterus) and haematological malignancies (such as leukaemia). While their lower positions on the causes of DALYs rankings indicates their overall incidence in the respective Sub-Saharan African region is considerably less than some other causes of DALYs, this does not necessarily mean that malignancies are irrelevant or less important. These conditions may not occur as frequently in comparison to the magnitude of other diseases in the Sub-Saharan African context, but they are important and graduates would be expected to recognise the manifestations of these disease conditions and know what would be the next most appropriate step to take even if they are unable to provide the appropriate health care themselves. This draws attention to a further limitation of the DALYs in that they would not be useful in the context of a symptom-based curriculum. The initial presentation of diseases is not always clear cut and arriving at a diagnosis requires careful differentiation of presenting symptoms and signs. This would have implications for curriculum planning and will be discussed further in Section 6.3.

6.2.1.4 Weighting of marks

One further point that deserves mention here concerns the weighting of marks between clinical clerkships and final examinations, indicated as a 40:60 ratio at Makerere and Stellenbosch Universities (see Tables 5.2 and 5.3). Reasons for the decisions behind these weightings are not evident but there is good sense to believe that students might interpret this to mean that what is tested during clinical clerkships is less important than what is assessed during exit examinations (Cilliers et al., 2010; Cilliers et al., 2012). Given that assessment generally featured at the lower tiers of Miller's 'pyramid' during clinical clerkships (See Section 5.2), this might result in a double-negative impact in that examiners have less useful information to base decisions on and students may ignore important competencies they will need in practice.

6.2.2 Examiners as Key Players

Following on from assessment in practice, attention is now turned to a second key issue that emerged from this study: the influence examiners have on exit-level assessment perhaps as key players involved in the assessment.

As has been demonstrated in Section 5.4.4, **examiners** were aware of the health-care needs in their respective local contexts and that the focus of their training needed to ensure that their students could cope with the demands they would face upon graduation as doctors. They had an understanding of clinical competence which essentially centred on the acquisition of basic skills and mastery of all the steps involved in clinical management. The examiners, however, recognised gaps in assessment, such as the testing of communication skills which currently was not emphasized. Patient safety was highlighted as a concern, yet there was scant evidence that this specifically informed decisions about the content or methods of assessment used. This draws attention to one of the limitations of using self-reported data, as reported by Hartman and Nelson (1992) in their study on medical school faculty, namely that there was a poor correlation between what was said (self-reported teaching behaviour) and what was done (as performance in written simulations). Nevertheless, the data gathered from the interviews in this study allowed an exploration, although limited, of discrepancies between the two sets of data (what they said in the interviews, and the observations made of the exit-level assessments), and this could be considered a strength of this study. Consideration could be given to exploring these discrepancies further in future research.

In the study interviews, examiners also expressed preconceived ideas about what should or should not be known or demonstrated in assessment. These assessment expectations differed between examiners, and also between institutions. In addition, the source of these expectations also differed. Some examiners, for example, based their expectations on their own personal experiences or from observations made during training they had received overseas (outside Sub-Saharan Africa) or from collaboration with institutions overseas (See Section 5.4.2). It is also plausible that these expectations could have arisen from their own experiences dealing with the spectrum of clinical conditions encountered in their daily clinical practice. These preconceived ideas bear similarities to the “intuitive beliefs” about summative assessment as described by

Harrison et al. (2017). Although the context of Harrison et al.'s study concerned the challenges of changing the culture of assessment whereas this study relates to the validity of exit-level assessment practices in determining the clinical competence of medical graduates, in both instances a better understanding of these ideas or beliefs will need to be achieved before any attempts to implement change can be considered. With such preconceived ideas on assessment, it is not unimaginable that examiner expectations could influence the way in which they apply rubrics and interpret checklists in assessment.

Interview participants were also aware that there were occasions when their conduct as examiners was not objective and they were open to bias. With regard to the influence of examiner behaviour on assessment practices, it is well recognised that the issue of variation in examiner behaviour is a major contributory factor to the poor reliability and validity of orals (Daelmans et al., 2001; Kearney et al., 2002; Wass et al., 2003; Davis and Karunathilake, 2005; Memon et al., 2010) and these would need to be addressed because of their impact on the performance of students. This behaviour invariably has subjective elements which could have arisen from personal experiences. It could also have come about as a result of their observing the way other examiners conduct themselves during assessments.

Examiners did not appear to be aware of assessment concepts such as blueprinting and standard setting (See Section 2.6). Some were mindful of the importance of ensuring their students had the required competence for clinical practice and that assessment methods should be aligned accordingly (See Section 5.4.1, B04Y120:26) but there was no evidence that these views were widely shared by other examiners. Some educational training and support on these issues could perhaps help heighten their awareness and strengthen their understanding of these issues. This feature will be revisited in Section 6.3.

In reviewing the findings that have emerged from this study about the key role examiners plays in assessment, it has become apparent that they do have a major influence on assessment practices. Their awareness and knowledge of the health-care demands of their location probably does influence their teaching, and by implication, could then impact on the assessment of their students in terms of choice of method(s)

as well as assessment content. This is an area that could benefit from further future research.

Thus it is evident examiners have a key role in exit-level assessment, and that this is only one of several character roles that they play depending on the situation that they are in. These individuals who assess students in examinations are typically also the same individuals who teach the students as well as provide clinical service to the community and they are therefore conversant with local health-care needs. Some overlap of these roles is probably inevitable and could influence what they do in their assessor role. They may in turn be influenced by the institution that they teach and assess in, and this leads the discussion to the next key issue which emerged from this study, of context and institutional culture.

A second point of note relates to the extent to which examiners have an understanding of basic concepts of assessment, termed "**assessment literacy**" in the literature (Stiggins, 1991; Stiggins, 1995; Popham, 2011). This has been defined as "an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (Popham, 2011:267). One example given earlier relates to the blueprinting and standard setting where examiners did not appear to be aware of these concepts. As another example (See Section 5.4.3.2), a number of respondents in the interviews were unable to provide any explanation or a direct answer to the query about reasons for choices of exit-level assessment methods. There could be several possible explanations for the tacit silence on this issue, one of which could be that examiners simply did not know the reasons for the choice of assessment methods and did not wish to admit this, or they may have accepted the choice of methods without questions but not understood the reasons for them. A second reason could be that there were institutional issues at play, and this will be discussed further in the following Section 6.2.3.

6.2.3 The influence of context and institutional culture

6.2.3.1 Context of assessment

The role of context emerged as a key factor influencing the assessment methods selected. It would appear that most of the in-course assessments taking place in the exit-level clinical clerkships were conducted in a clinical setting, either ward- or clinic-

based. In comparison, only three out of the six study situations conducted final clinical examinations in an actual clinical setting (See Section 5.1 and Table 5.1). Making use of existing available opportunities and resources especially with regard to clinical material lends weight to the authenticity of the assessment in terms of location and clinical context. Clinical cases selected from real patients (as inpatients admitted to the ward for treatment or as outpatients coming to clinic) usually depend on the patients that are available at the time the examination is being conducted. This could limit the spectrum of conditions available leading to students being assessed on a limited number of conditions. A further limitation would be that patients who are seriously ill would be excluded from the pool of patients available for student examinations.

The use of real patients and real clinic settings may be less costly than having purpose-built facilities to simulate wards or clinics with simulated patients, and would help provide a chance to better assess what the student is likely to do in actual practice (Collins and Harden, 1998; Schuwirth and Van der Vleuten, 2003b). However, using actual wards or clinics for conducting examinations can lead to disruption of routine clinical services and may not always be feasible. This needs to be taken into consideration together with other logistic issues such as the numbers of students and patients involved in such an exercise. Current available literature on creating authentic clinical settings for assessment tends to focus more on the use of patients (real or simulated) to create as real a setting as possible, with less emphasis on actual venues used.

Institutions also provide a context in that they form the space within which assessment is enacted. This will be incorporated into the discussion regarding institutional culture which is to follow.

6.2.3.2 Institutional culture

Institutions could also be regarded as key players in influencing assessment practices. There are institutional policies, practices and norms that are dominant in the institutional space that could influence assessment practices. Some examiners seemingly responded to norms (unwritten rules that are expected or considered as standard) when perpetuating a particular assessment practice (See Section 5.4.3). Questions arise about the role played by past institutional practices in the understanding of current assessment practices, and the extent to which policy decisions

made at institutional management level are conveyed to faculty who are responsible for interpreting policy and executing the assessments themselves. In the previous section 6.2.2, attention was drawn to the fact that a number of interview participants were unable to provide an explanation for the choices of exit-level assessment methods used in their respective institution. Some of the more senior and experienced faculty did know the reasons and related the choice of methods back to institutional policy decisions that were determined some time previously, and more recently appointed faculty may not have been aware of these decisions. This suggests there may be areas in which faculty development would be required. However, an investigation of existing institutional policies and practices would first help determine the current status in order to help forward planning.

6.3 Study Implications

In considering the broader context of Sub-Saharan Africa, other medical schools on the continent may well demonstrate similarities to the assessment practices that have been found in the three medical schools that were investigated in this study. If the challenge of dealing with resource constraints is a common shared finding in the majority of Sub-Saharan African countries, then it would have greater relevance to consider exploring mechanisms to enable the sharing of information with regard to practices, strategies and solutions.

Having provided a synthesis of the key findings, it is necessary to consider the implications thereof, particularly in light of the research questions that were posed for this study. In this section, these implications are discussed. Some recommendations for further research and/or action are also offered.

6.3.1 Validity of exit-level assessment

The study findings suggest that the exit-level assessment taking place in two clinical disciplines in three selected Sub-Saharan African medical schools may have limited validity (See Section 6.2.1.1). The implications here relate to issues of patient safety and licensing for medical practice which were described in Chapter One (Section 1.1) as being central to clinical competence of graduates from medical schools. Doctors need to be able to handle the high burden of disease seen in Sub-Saharan Africa and have the necessary knowledge and skills required for the clinical management of disease

conditions. Their quality of health-care provided will be affected if they are ill-prepared and unable to cope with the health-care needs of the community that they are practising in. This situation would raise concerns about patient safety, which regulatory and licensing authorities would need to take seriously.

It can be argued that a better alternative for making decisions about the clinical competence of students at the end of their medical training could be by adopting some form of workplace-based assessment (Norcini, 2003a; Norcini and Burch, 2007) as an alternative to the current summative assessment practices taking place, and would be a potential area for further research.

6.3.2 The importance of alignment

As indicated earlier in the study, at the level of the curriculum, the teaching and assessment provided has to be aligned in order to provide education that has relevance in application. Biggs' Theory of Constructive Alignment (1996) speaks to the interconnectedness between these three dimensions (outcome-teaching-assessment) and this desired alignment. The earlier discussion can be reviewed in light of this theoretical position and highlights questions that curriculum developers and assessment planners in medical schools with similar backgrounds may wish to consider. While one would seek to ensure that there is alignment between the curriculum and the assessment (content and approach), it is equally critical that there is alignment between context and curriculum. If the latter is missing, then the former can be seen to be less relevant or less useful. Thus, in the context of this study, alignment has emerged as a key construct at a more meta-level as well. Frenk et al.'s (2010) alignment between health system and educational system speaks to the implications of this study at a systemic level, and in so doing, adds another layer to the complexity of this space. It would make sense for these two groupings also to be aligned to each other, since they are both involved in the provision of medical graduates to meet the local health-care needs of the communities in which they serve. All of this has implications for the way in which assessment is conceived, designed and implemented, and for the validity of judgments being made about the competence of medical graduates.

6.3.3 Curriculum reform/revision

While it may seem straightforward to say that teaching and assessment in the curriculum should be aligned to contextual needs, this study has shown the difficulties encountered in tackling a revision of such an alignment. Perhaps a good starting point would be institutional review of the curriculum, with careful analysis of any existing problems by drawing upon institutional quality assurance processes, external reviews (such as from national licensing authorities, educational councils, external examiner reports) and wide consultations with a range of stakeholders (Davis and Harden, 2003; Bordage and Harris, 2011). With regard to assessment in particular, reviews provided by external examiners could help identify the strengths and weaknesses of assessment systems, which could then impact on curriculum development (Morrison and Watt, 2002; Quality Assurance Agency for Higher Education UK, 2017).

Findings from this study have indicated how important a fine-grained evaluation of exit-level assessment methods is as part of any form of revision. The findings of this study suggest a need to shift assessment from the lower tiers of Miller's 'pyramid' ("Knows" and "Knows How") to the higher "Shows How" and "Does" tiers. As an example, several of the OSPE stations in the Obstetrics & Gynaecology OSCE/OSPE at the University of Stellenbosch could be modified to be at the "Shows How" tier. This poses new questions: What proportion of assessment should ideally be at any given level of Miller's 'pyramid' at exit-level? What would be considered as "adequate" in the assessment of clinical competence? What mix of assessment methods should medical schools in Sub-Saharan Africa be striving for? Should regulators be changing their expectations? The literature is silent on this beyond a generic exhortation to assess competence at exit-level. Having some sense of what mix of knowledge and competence assessment is appropriate could then serve as a starting point for a review of exit-level assessment methods and to guide further research in this area.

Taking cognisance of the resource constraints and local context in Sub-Saharan Africa, it would be worth considering an integrated approach and whether this would allow more meaningful decisions to be made about the competence of medical graduates. This has already been raised earlier (in Section 6.2.1.2). In tandem with the integrated teaching and learning activities, integrated assessment would also need to be considered. There are implications in terms of manpower, such as whether integrated assessment would

make use of fewer examiners. Examination questions used in integrated assessment would also need to be constructed by faculty representing different contributing disciplines (Malik and Malik, 2011:103), and this may pose challenges that need to be addressed during the curriculum review. Faculty are already evidently stretched to meet the varying and extensive demands made of them. Adding to this another set of meetings to design assessment will be a change strategy that needs thorough planning prior to implementation.

With curriculum renewal, the issues of making change and overcoming resistance to change need to be considered. Harrison et al. have made some pertinent comments concerning the challenges of changing the culture of assessment. Students and faculty would need to change how they conceptualise assessment, and this would mean they "need to stop believing in the primacy of summative high-stakes assessment" (Harrison et al., 2017:12). There would also be implications here for faculty development, and this will be discussed later in Section 6.3.6.

6.3.4 Use of DALYs in blueprinting

The study findings show that there is much that is relevant in the content of curricula in the study sites, but that there is also content that is called into question regarding its contextual relevance. Mapping the content of a given examination would provide a way of ensuring adequate relevance and sampling. This process, known as blueprinting (Section 2.6), involves the development of a template used to help map content against learning objectives. Disability Adjusted Life Years (DALYs) have the potential to contribute to a process of blueprinting (along with other parameters or tools), to ensure an appropriate breadth of sampling of curriculum topics in assessment. This would have an important function in strengthening the validity of the assessment content (Hamdy, 2006; Coderre et al., 2009). As the DALYs lists are not fixed, and vary across WHO regions as well as over time, this could be considered a strong point as it would require ongoing review of curricula and thereby heighten currency and relevance. However, the use of DALYs in this study has also revealed limitations such as the focus on disease conditions only, and such limitations may also manifest when attempts are made to use DALYs as a guide for blueprint construction. This study focussed on two clinical disciplines only, and therefore not all causes of DALYs applied to these disciplines. As noted above, DALYs also focus on disease but are silent on health

promotion, prevention, a primary health care approach. The findings of this study illustrate how DALYs would have to be supplemented by another framework to ensure balance and overall relevance during curriculum design. One such consideration as an additional strategy could be an evaluation of whether physician intervention could make a significant difference to the outcomes of patients.

6.3.5 Defining institutional guidelines and policies

Institutions have a responsibility to provide information on institutional policies, standards and procedures as reference points to guide faculty, and faculty have a concomitant responsibility to critically engage with their practice and remain abreast of policy. In the course of this study, institutional policies (for example, policies on assessment practices) were not available for perusal and therefore faculty would place at a disadvantage.

6.3.6 Faculty Development

The findings of this study are not conclusive in terms of the extent to which faculty development is indicated, but they do offer tantalising evidence that assessment literacy is limited – a limitation that could be addressed through meaningful conversations around the principles and practice of assessment. Some findings suggest that examiners do not understand the practices they are part of, nor do they exercise agency to change practices that are suboptimal. This points to an opportunity for faculty development programmes to provide education training and assist health-care professionals in their role as effective teachers. Faculty development activities in health professions education are not new and this field has been extensively researched over the past 40 years (Wilkerson and Irby, 1998; Harden and Crosby, 2000; Challis, 2001; Steinert and Mann, 2006; Higgs and McAllister, 2007; McLean et al., 2008; Swanwick, 2008; Steinert et al., 2012). Part of the focus of faculty development would also need to be on providing some grounding in principles of medical education, as well as aspects of curriculum and assessment relevant to their needs. This would enable a clear understanding of the components that contribute to the desired constructive alignment described earlier. The setting up of a faculty development programme (if there is none already in place) would require extensive groundwork and planning subject to available resources, as well as strong support from institutional management. However, such a

plan for faculty development could create a dilemma in settings where there are already severe resource constraints, and although ways of making use of existing resources and personnel will need to be included in the planning, this may be insufficient to deal with such a problem.

6.4 Lessons learnt from this study

Much has been learnt from this study with insights gained into exit-level assessment practices. Although these three selected medical schools studied are individual cases, they provide just one set of perspectives that cannot be seen to represent Sub-Saharan Africa; it is important to consider the lessons learnt that could have relevance for other medical schools in similar contexts.

A common finding in this study has been that at exit-level a range of assessment methods are being used. In mapping these against an analytical framework that provides for different levels of engagement by the students, it became evident that the range of methods does not necessarily mirror the range of levels along the analytical frame. In this context, the mix of assessment methods may be of less value than initially thought. Importantly, though, mapping the content of assessment against a second analytical instrument indicated that the content had some relevance to the local health context.

However, the findings from this study have also shown that the exit-level assessment taking place in two clinical disciplines in these three selected medical schools does not fully allow us to draw valid inferences that medical graduates have the clinical competence to meet the demands of practice in their respective settings.

Faculty play a prominent role in medical schools in teaching students and in enacting the assessment that takes place in the medical programme, although they do not appear to have much influence on the design of assessment. In order to consider addressing any changes to assessment practices, examiners will need to be engaged first over any preconceptions they have on assessment as well as issues of academic literacy.

While it would be simple to attribute the state of current suboptimal assessment practices in a Sub-Saharan African setting to well-recognised and existing resource constraints, this study has also indicated that it is plausible for the assessment literacy of examiners to be a contributing factor. There is insufficient firm evidence from this

study to support either suggestion, but this would be suitable ground for future research.

6.5 Study Limitations

The study took place in three selected medical schools in the continent where English is a medium of instruction. Medical schools in Francophone, Lusophone and Arabic-speaking countries, which make up 60% of the countries in Sub-Saharan Africa, were not studied, and it is not known whether the same assessment methods and practices would be used in undergraduate medical training programmes in these countries, or whether the institutional norms or culture that were seen to play an important part in influencing assessment practices in the medical schools that were studied would apply equally in other countries on the continent. These are limitations in terms of the potential of the study to offer generalisations to other Sub-Saharan African medical schools. In addition, the search for relevant literature relating to exit-level assessment and the Sub-Saharan African context was conducted in English language journals and publications. This means that literature or information from sources in other languages may not have been included in the review, and this is a further limitation.

Examining several case studies allows an analysis within each setting and across settings to understand the similarities and differences that exist (Baxter and Jack, 2008:550). Although the evidence gathered is considered robust and reliable, multiple case studies are time consuming and expensive to conduct (Yin, 2014:57). The evidence gathered at the three selected medical schools in this study related to the exit-level assessment practices in two clinical disciplines – Obstetrics & Gynaecology and Paediatrics – thus focussing on three clearly defined cases. This was a deliberate decision made during the design phase of the study to ensure the study was feasible.

While using DALYs as an analytical framework was shown to be immensely useful, as described in Section 5.3.2 in the initial phase of data analysis, and again later in Section 6.1.1.2, the DALYs proved to have limitations as an analytical framework.

One of the challenges during data collection related to the recruitment of faculty for interviews and this meant that recruitment did not follow a consistent process but varied considerably between the three study sites and in the two disciplines, as already described in Section 4.4.3.1. In hindsight, it would have been ideal to adhere to using

the same approach to participant recruitment but this was not possible to follow since the circumstances surrounding each situation were different, as would be expected when using multiple cases. These experiences emphasized how tacit approval of the respective head of department for the study to be conducted was crucial, even if clinical clerkship coordinators or individual faculty were agreeable and willing to participate in the study. Nevertheless, this approach yielded data that was worthwhile as study participants did express their views freely and willingly. The criteria governing the rigour of data collection and the trustworthiness of the data collected (See Section 4.4.3.1) were not compromised.

Data saturation may not have been attained in this study as the interview participants who responded to the study invitation were drawn from the fixed population of faculty involved in exit-level assessment in their respective discipline. It is a challenge to recruit sufficient numbers of participants when the potential study population is already small in number, and other than endeavouring to ensure a better recruitment rate, this remains a challenge for work in this context. Where data collection involves expensive travel to and sojourn in another country, there is a defined time period within which data can be collected, with little or no room to manoeuvre should respondents not be available.

There were some instances of missing data, such as the model answers for examination questions. Effort was made to obtain the missing information on request, but these attempts were not uniformly successful. Possible explanations for this include the challenges of communicating with busy clinicians and also the limited reach of electronic communication channels. It could be that they may not have had model answers prepared or may not have been willing to release the model answers. However, these explanations are purely speculative. In these instances of missing data, the clarificatory interviews with clinical clerkship coordinators or heads of department were used to try and supplement the information that was available from document analysis and observations.

This study only focussed on two clinical disciplines. However, they are two quite distinct disciplines which cannot be considered as similar, with Obstetrics & Gynaecology being a more surgical-based discipline and focussing on one population group (women of child-bearing age), whereas Paediatrics is a more medical-based

discipline and focusses on a different and younger population group (children). A next step would be to extend the study to focus on other final-year clinical disciplines in the same institutions to investigate the extent to which those findings would be comparable to the ones from this study.

6.6 Conclusions

As a reminder of the research question, this study set out to investigate the validity of exit-level assessment practices in selected medical schools in Sub-Saharan Africa in determining the clinical competence of medical graduates for selected health issues in this resource-constrained context. What I found was that although assessment question content was appropriate and focussed on relevant contextual health-related issues, current assessment practices did not provide sufficient opportunity for data to be gathered and for decisions to be made about competence.

Disability Adjusted Life Years (DALYs) have the potential to contribute to a process of blueprinting to ensure an appropriate breadth of sampling of curriculum topics in assessment. Faculty as examiners as well as institutional culture and practices all have significant influence on assessment practices that take place in an institution. Faculty development could be considered in addressing issues such as assessment literacy. Alignment of teaching, assessment and outcomes of a curriculum is important, as is the alignment between the education and health systems, with these two entities working in conjunction with each other.

Finally, I would like to end this study and the journey I embarked on five years ago with some personal reflections. What have been apparent to me throughout the course of this study are the continuing significant challenges faced by many countries in Sub-Saharan Africa. This study has afforded me an opportunity to see and experience some of these challenges first hand, and it has served as a reminder of the role I can play as a practising clinician and medical educator in striving to make the teaching and learning of medicine more relevant and valid. What I have learnt has already had an impact on the way I teach, reflect and give feedback to my students, and most certainly in the way I assess them. I trust this will be to our mutual benefit.

> "*Give a man a fish and you feed him for a day; teach a man to fish and you feed him for a lifetime*" (Source unknown).

References

ACADEMY OF MEDICAL ROYAL COLLEGES. 2009. *Improving assessment* [Online]. London: Academy of Medical Royal Colleges Press. Available: http://www.aomrc.org.uk/publications/reports-a-guidance?view=docman [Accessed 28 April 2015].

AFRICAN JOURNALS ONLINE. 2015. *African Journals Online (AJOL)* [Online]. Grahamstown: AJOL. Available: http://www.ajol.info/ [Accessed 02 February 2015].

ALBANESE, M. A., MEJICANO, G., MULLAN, P., KOKOTAILO, P. & GRUPPEN, L. 2008. Defining characteristics of educational competencies. *Medical Education,* 42**,** 248-255.

ALBERT, M. 2004. Understanding the debate on medical education research: A sociological perspective. *Academic Medicine,* 79**,** 948-954.

AMEE. 2015. *Association for Medical Education in Europe* [Online]. Dundee: Association for Medical Education in Europe. Available: http://www.amee.org/what-is-amee [Accessed 13 January 2015].

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION & JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL PSYCHOLOGICAL TESTING 1999. *Standards for educational and psychological testing,* American Educational Research Association.

ANDERSON, J. 2004. Multiple-choice questions revisited. *Medical Teacher,* 26**,** 110-113.

ANDERSON, M., COHEN, J., HALLOCK, J., KASSEBAUM, D., TURNBULL, J. & WHITCOMB, M. 1999. Learning objectives for medical student education: Guidelines for medical schools: Report I from the Medical School Objectives Project. *Academic Medicine,* 74**,** 13-18.

ANYANGWE, S. C. & MTONGA, C. 2007. Inequities in the global health workforce: The greatest impediment to health in sub-Saharan Africa. *International Journal of Environmental Research and Public Health,* 4**,** 93-100.

ASTOR, A., AKHTAR, T., MATALLANA, M. A., MUTHUSWAMY, V., OLOWU, F. A., TALLO, V. & LIE, R. K. 2005. Physician migration: views from professionals in Colombia, Nigeria, India, Pakistan and the Philippines. *Social Science & Medicine,* 61**,** 2492-2500.

AUSTRALIAN MEDICAL COUNCIL. 2012. *Standards for assessment and accreditation of primary medical programs by the Australian Medical Council* [Online]. Kingston ACT: Australian Medical Council. Available: http://www.amc.org.au/joomla-files/images/Accreditation/FINAL-Standards-and-Graduate-Outcome-Statements-20-December-2012.pdf [Accessed 24 January 2014].

AVELING, E.-L., KAYONGA, Y., NEGA, A. & DIXON-WOODS, M. 2015. Why is patient safety so hard in low-income countries? A qualitative study of healthcare workers' views in two African hospitals. *Globalization and Health,* 11**,** 6.

BABBIE, E. R. & MOUTON, J. 2001. *The practice of social research,* Cape Town, Oxford University Press Southern Africa.

BANDARANAYAKE, R. C. 2008. Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Medical Teacher,* 30**,** 836-845.

BARZANSKY, B. & ETZEL, S. I. 2003. Educational programs in US medical schools, 2002-2003. *JAMA,* 290**,** 1190-1196.

BAXTER, P. & JACK, S. 2008. Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report,* 13**,** 544-559.

BIGGS, J. 1996. Enhancing teaching through constructive alignment. *Higher Education,* 32**,** 347-364.
BIGGS, J. 2014. Constructive alignment in university teaching. *HERDSA Review of Higher Education,* 1**,** 5-22.
BIGLAN, A. 1973. The characteristics of subject matter in different academic areas. *Journal of Applied Psychology,* 57**,** 195-203.
BLOOM, B. S. 1971. *Handbook on formative and summative evaluation of student learning,* New York, McGraw-Hill.
BORDAGE, G. 2009. Conceptual frameworks to illuminate and magnify. *Medical Education,* 43**,** 312-319.
BORDAGE, G. & HARRIS, I. 2011. Making a difference in curriculum reform and decision-making processes. *Medical Education,* 45**,** 87-94.
BOUD, D. 2000. Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education,* 22**,** 151-167.
BOUD, D. & FALCHIKOV, N. 2006. Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education,* 31**,** 399-413.
BOULET, J., BEDE, C., MCKINLEY, D. & NORCINI, J. 2007. An overview of the world's medical schools. *Medical Teacher,* 29**,** 20-26.
BOURSICOT, K., ETHERIDGE, L., SETNA, Z., STURROCK, A., KER, J., SMEE, S. & SAMBANDAM, E. 2011. Performance in assessment: Consensus statement and recommendations from the Ottawa conference. *Medical Teacher,* 33**,** 370-383.
BOURSICOT, K. & ROBERTS, T. 2006. Setting Standards in a Professional Higher Education Course: Defining the Concept of the Minimally Competent Student in Performance-Based Assessment at the Level of Graduation from Medical School. *Higher Education Quarterly,* 60**,** 74-90.
BOWEN, G. A. 2009. Document analysis as a qualitative research method. *Qualitative Research Journal,* 9**,** 27-40.
BRAUER, D. G. & FERGUSON, K. J. 2015. The integrated curriculum in medical education: AMEE Guide No. 96. *Medical Teacher,* 37**,** 312-322.
BRIGHTWELL, A. & GRANT, J. 2013. Competency-based training: who benefits? *Postgraduate Medical Journal,* 89**,** 107-110.
BRITANNICA, E. I. 2017. Continent. Chicago, IL: Encyclopædia Britannica.
BRITTEN, N. 1995. Qualitative interviews in medical research. *British Medical Journal,* 311**,** 251-253.
BROOKS, M. A. 2009. Medical education and the tyranny of competency. *Perspectives in Biology and Medicine,* 52**,** 90-102.
BROWN, C., ROSS, S., CLELAND, J. & WALSH, K. 2015. Money makes the (medical assessment) world go round: The cost of components of a summative final year Objective Structured Clinical Examination (OSCE). *Medical Teacher,* 37**,** 653-659.
BUCHAN, J., BALDWIN, S. & MUNRO, M. 2008. Migration of health workers: the UK perspective to 2006. *OECD Health Working Papers* [Online]. Available: http://dx.doi.org/10.1787/228550573624 [Accessed 09 April 2015].
BUNNISS, S. & KELLY, D. R. 2010. Research paradigms in medical education research. *Medical Education,* 44**,** 358-366.
BURCH, V. C. & SEGGIE, J. L. 2008. Use of a structured interview to assess portfolio-based learning. *Medical Education,* 42**,** 894-900.
CAMBRIDGE UNIVERSITY PRESS 2013. Cambridge Advanced Learner's Dictionary. 4th ed.: Cambridge University Press.
CAMERON, D., WOLVAARDT, L., VAN ROOYEN, M., HUGO, J., BLITZ, J. & BERGH, A.-M. 2011. Medical student participation in community-based experiential learning:

reflections from first exposure to making the diagnosis. *South African Family Practice,* 53**,** 373-379.
CARRACCIO, C. & BURKE, A. E. 2010. Beyond competencies and milestones: adding meaning through context. *Journal of Graduate Medical Education,* 2**,** 419-422.
CARRACCIO, C., WOLFSTHAL, S. D., ENGLANDER, R., FERENTZ, K. & MARTIN, C. 2002. Shifting paradigms: from Flexner to competencies. *Academic Medicine,* 77**,** 361-367.
CASE, S. & SWANSON, D. 2002. Constructing written test questions from the basic and clinical sciences. 3rd ed. Philadelphia: National Board of Medical Examiners.
CELLETTI, F., REYNOLDS, T., WRIGHT, A., STOERTZ, A. & DAYRIT, M. 2011. Educating a new generation of doctors to improve the health of populations in low- and middle-income countries. *PLoS Medicine* [Online], 8. [Accessed 18 February 2015].
CHALLIS, M. 2001. Building an effective programme for clinical teachers: the role of the staff developer. *Medical Teacher,* 23**,** 270-275.
CHAMBERS 2014. The Chambers Dictionary 13th ed.: Chambers.
CHEN, C., BUCH, E., WASSERMANN, T., FREHYWOT, S., MULLAN, F., OMASWA, F., GREYSEN, S. R., KOLARS, J. C., DOVLO, D., BAKR, D. E. E. G. A., HAILEAMLAK, A., KOUMARE, A. K. & OLAPADE-OLAOPA, E. O. 2012. A survey of Sub-Saharan African medical schools. *Human Resources for Health* [Online], 10. Available: http://www.human-resources-health.com/content/10/1/4 [Accessed 09 October 2014].
CHEN, L., EVANS, T., ANAND, S., BOUFFORD, J. I., BROWN, H., CHOWDHURY, M., CUETO, M., DARE, L., DUSSAULT, G., ELZINGA, G., FEE, E., HABTE, D., HANVORAVONGCHAI, P., JACOBS, M., KUROWSKI, C., MICHAEL, S., PABLOS-MENDEZ, A., SEWANKAMBO, N., SOLIMANO, G., STILWELL, B., WAAL, A. D. & WIBULPOLPRASERT, S. 2004. Human resources for health: overcoming the crisis. *The Lancet,* 364**,** 1984-1990.
CILLIERS, F. J., SCHUWIRTH, L. W., ADENDORFF, H. J., HERMAN, N. & VAN DER VLEUTEN, C. P. 2010. The mechanism of impact of summative assessment on medical students' learning. *Advances in Health Sciences Education,* 15**,** 695-715.
CILLIERS, F. J., SCHUWIRTH, L. W., HERMAN, N., ADENDORFF, H. J. & VAN DER VLEUTEN, C. P. 2012. A model of the pre-assessment learning effects of summative assessment in medical education. *Advances in Health Sciences Education,* 17**,** 39-53.
CLELAND, J., SCOTT, N., HARRILD, K. & MOFFAT, M. 2013. Using databases in medical education research: AMEE Guide No. 77. *Medical Teacher,* 35**,** e1103-e1122.
CLELAND, J. A., ABE, K. & RETHANS, J.-J. 2009. The use of simulated patients in medical education: AMEE Guide No 42. *Medical Teacher,* 31**,** 477-486.
CODERRE, S., WOLOSCHUK, W. & MCLAUGHLIN, K. 2009. Twelve tips for blueprinting. *Medical Teacher,* 31**,** 322-324.
COLLINS 2017. Collins English Dictionary. Collins.
COLLINS, J. & HARDEN, R. 1998. AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. *Medical Teacher,* 20**,** 508-521.
COOK, D. A., BORDAGE, G. & SCHMIDT, H. G. 2008. Description, justification and clarification: a framework for classifying the purposes of research in medical education. *Medical Education,* 42**,** 128-133.

COWAN, D. T., NORMAN, I. & COOPAMAH, V. P. 2005. Competence in nursing practice: A controversial concept – A focused review of literature. *Nurse Education Today,* 25**,** 355-362.

CRESWELL, J. W. 2014. *Research design: Qualitative, quantitative, and mixed methods approaches,* Thousand Oaks, CA, Sage.

CROSS, M., MUNGADI, R. & ROUHANI, S. 2002. From policy to practice: Curriculum reform in South African education. *Comparative Education,* 38**,** 171-187.

CROSSLEY, J., HUMPHRIS, G. & JOLLY, B. 2002. Assessing health professionals. *Medical Education,* 36**,** 800-804.

CROSSLEY, J. & JOLLY, B. 2012. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Medical Education,* 46**,** 28-37.

CROWE, S., CRESSWELL, K., ROBERTSON, A., HUBY, G., AVERY, A. & SHEIKH, A. 2011. The case study approach. *BMC Medical Research Methodology,* 11**,** 100.

CRUESS, R., MCILROY, J. H., CRUESS, S., GINSBURG, S. & STEINERT, Y. 2006. The Professionalism Mini-Evaluation Exercise: A preliminary investigation. *Academic Medicine,* 81**,** S74-S78.

CRUESS, R. L., CRUESS, S. R. & STEINERT, Y. 2016. Amending Miller's pyramid to include professional identity formation. *Academic Medicine,* 91**,** 180-185.

CUMMING, A. & ROSS, M. 2007. The Tuning Project for Medicine – learning outcomes for undergraduate medical education in Europe. *Medical Teacher,* 29**,** 636-641.

CUSIMANO, M. D., COHEN, R., TUCKER, W., MURNAGHAN, J., KODAMA, R. & REZNICK, R. 1994. A comparative analysis of the costs of administration of an OSCE (Objective Structured Clinical Examination). *Academic Medicine,* 69**,** 571-6.

DAELMANS, H. E., SCHERPBIER, A. J., VAN DER VLEUTEN, C. P. & DONKER, A. J. 2001. Reliability of clinical oral examinations re-examined. *Medical Teacher,* 23**,** 422-424.

DANNEFER, E. F. 2013. Beyond assessment of learning toward assessment for learning: Educating tomorrow's physicians. *Medical Teacher,* 35**,** 560-563.

DANZON, P. M. 1990. The Crisis in Medical Malpractice: A Comparison of Trends in the United States, Canada, the United Kingdom and Australia. *Law, Medicine & Health Care,* 18**,** 48.

DAVIES, P. 1999. What is evidence-based education? *British Journal of Educational Studies,* 47**,** 108-121.

DAVIS, M., BEN-DAVID, M. F., HARDEN, R., HOWIE, P., KER, J., MCGHEE, C., PIPPARD, M. & SNADDEN, D. 2001. Portfolio assessment in medical students' final examinations. *Medical Teacher,* 23**,** 357-366.

DAVIS, M. H. 2003. OSCE: the Dundee experience. *Medical Teacher,* 25**,** 255-261.

DAVIS, M. H. & HARDEN, R. M. 2003. Planning and implementing an undergraduate medical curriculum: the lessons learned. *Medical Teacher,* 25**,** 596-608.

DAVIS, M. H. & KARUNATHILAKE, I. 2005. The place of the oral examination in today's assessment systems. *Medical Teacher,* 27**,** 294-297.

DE CHAMPLAIN, A. F. 2004. Ensuring that the competent are truly competent: an overview of common methods and procedures used to set standards on high-stakes examinations. *Journal of Veterinary Medical Education,* 31**,** 62-66.

DENSCOMBE, M. 2010. *The Good Research Guide for Small-Scale Social Research Projects,* Maidenhead, OUP in association with McGraw-Hill Education.

DEPARTMENT OF EDUCATION (DOE) 1997. Education White Paper 3: A programme for the transformation of higher education. *In:* EDUCATION, D. O. (ed.). Pretoria: Department of Education.

DICICCO-BLOOM, B. & CRABTREE, B. F. 2006. The qualitative research interview. *Medical Education,* 40**,** 314-321.
DOWNING, S. M. 2003. Validity: on the meaningful interpretation of assessment data. *Medical Education,* 37**,** 830-837.
DOWNING, S. M. 2004. Reliability: on the reproducibility of assessment data. *Medical Education,* 38**,** 1006-1012.
DOWNING, S. M., TEKIAN, A. & YUDKOWSKY, R. 2006. Research methodology: Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teaching and Learning in Medicine,* 18**,** 50-57.
DOWNING, S. M. & YUDKOWSKY, R. 2009. *Assessment in health professions education,* Routledge.
ELO, S. & KYNGÄS, H. 2008. The qualitative content analysis process. *Journal of Advanced Nursing,* 62**,** 107-115.
EPSTEIN, R. M. 2007. Assessment in medical education. *New England Journal of Medicine,* 356**,** 387-396.
EPSTEIN, R. M. & HUNDERT, E. M. 2002. Defining and assessing professional competence. *JAMA,* 287**,** 226-235.
EVA, K. W. & LINGARD, L. 2008. What's next? A guiding question for educators engaged in educational research. *Medical Education,* 42**,** 752-754.
FAIMER. 2011. *International Medical Education Directory* [Online]. Foundation for the Advancement of International Medical Education and Research. Available: http://www.faimer.org/resources/imed.html [Accessed 09 October 2014].
FERNANDEZ, N., DORY, V., STE-MARIE, L.-G., CHAPUT, M., CHARLIN, B. & BOUCHER, A. 2012. Varying conceptions of competence: an analysis of how health sciences educators define competence. *Medical Education,* 46**,** 357-365.
FOUAD, N. A., GRUS, C. L., HATCHER, R. L., KASLOW, N. J., HUTCHINGS, P. S., MADSON, M. B., COLLINS JR, F. L. & CROSSMAN, R. E. 2009. Competency benchmarks: A model for understanding and measuring competence in professional psychology across training levels. *Training and Education in Professional Psychology,* 3**,** S5-S26.
FOWELL, S., MAUDSLEY, G., MAGUIRE, P., LEINSTER, S. & BLIGH, J. 2000. Student assessment in undergraduate medical education in the United Kingdom, 1998. *Medical Education,* 34**,** 1-49.
FOWELL, S. L. & BLIGH, J. G. 1998. Recent developments in assessing medical students. *Postgraduate Medical Journal,* 74**,** 18-24.
FRAMBACH, J. M., MANUEL, B. A., FUMO, A. M., VAN DER VLEUTEN, C. P. & DRIESSEN, E. W. 2015. Students' and junior doctors' preparedness for the reality of practice in sub-Saharan Africa. *Medical Teacher,* 37**,** 64-73.
FRAMBACH, J. M., VAN DER VLEUTEN, C. P. & DURNING, S. J. 2013. AM Last Page: Quality Criteria in Qualitative and Quantitative Research. *Academic Medicine,* 88**,** 552.
FRANK, J. R. & DANOFF, D. 2007. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Medical Teacher,* 29**,** 642-647.
FRANK, J. R., MUNGROO, R., AHMAD, Y., WANG, M., DE ROSSI, S. & HORSLEY, T. 2010a. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Medical Teacher,* 32**,** 631-637.
FRANK, J. R., SNELL, L. S., CATE, O. T., HOLMBOE, E. S., CARRACCIO, C., SWING, S. R., HARRIS, P., GLASGOW, N. J., CAMPBELL, C. & DATH, D. 2010b.

Competency-based medical education: theory to practice. *Medical Teacher,* 32**,** 638-645.

FRENK, J., CHEN, L., BHUTTA, Z. A., COHEN, J., CRISP, N., EVANS, T., FINEBERG, H., GARCIA, P., KE, Y., KELLEY, P., KISTNASAMY, B., MELEIS, A., NAYLOR, D., PABLOS-MENDEZ, A., REDDY, S., SCRIMSHAW, S., SEPULVEDA, J., SERWADDA, D. & ZURAYK, H. 2010. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The Lancet,* 376**,** 1923-1958.

FRIEDMAN BEN-DAVID, M. 2013. Principles of Assessment. *In:* DENT, J. & HARDEN, R. M. (eds.) *A practical guide for medical teachers.* 4th ed. Edinburgh: Elsevier Health Sciences.

GAGE, N. L. 1989. The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher,* 18**,** 4-10.

GENERAL MEDICAL COUNCIL. 2009. *Tomorrow's Doctors: Recommendations on Undergraduate Medical Education* [Online]. London: General Medical Council UK. Available: http://www.gmc-uk.org/education/undergraduate/tomorrows_doctors_2009.asp [Accessed 20 February 2013].

GENERAL MEDICAL COUNCIL. 2011. *Assessment in undergraduate medical education. Advice supplementary to Tomorrow's Doctors (2009).* [Online]. London: General Medical Council UK. Available: http://www.gmc-uk.org/static/documents/content/Assessment_in_undergraduate-web.pdf [Accessed 20 February 2013].

GIBBS, G. R. 2007. *Analysing qualitative data,* Thousand Oaks, CA, Sage.

GINSBURG, S., REGEHR, G., HATALA, R., MCNAUGHTON, N., FROHNA, A., HODGES, B., LINGARD, L. & STERN, D. 2000. Context, conflict, and resolution: a new conceptual framework for evaluating professionalism. *Academic Medicine,* 75**,** S6-S11.

GITOMER, D. H. & ZISK, R. C. 2015. Knowing what teachers know. *Review of Research in Education,* 39**,** 1-53.

GLEESON, F. 1997. AMEE Medical Education Guide No. 9. Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). *Medical Teacher,* 19**,** 7-14.

GOLD, M. R., STEVENSON, D. & FRYBACK, D. G. 2002. HALYS and QALYS and DALYS, Oh My: Similarities and differences in summary measures of population Health. *Annual Review of Public Health,* 23**,** 115-134.

GOVAERTS, M. J. 2008. Educational competencies or education for professional competence? *Medical Education,* 42**,** 234-236.

GRANT, J. 1999. The incapacitating effects of competence: A critique. *Advances in Health Sciences Education,* 4**,** 271-277.

GREYSEN, S. R., DOVLO, D., OLAPADE-OLAOPA, E. O., JACOBS, M., SEWANKAMBO, N. & MULLAN, F. 2011. Medical education in sub-Saharan Africa: a literature review. *Medical Education,* 45**,** 973-986.

GUBA, E. G. 1981. Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ,* 29**,** 75-91.

GUKAS, I. D. 2007. Global paradigm shift in medical education: issues of concern for Africa. *Medical Teacher,* 29**,** 887-892.

HAGER, P. & GONCZI, A. 1996. What is competence? *Medical Teacher,* 18**,** 15-18.

HAGOPIAN, A., THOMPSON, M. J., FORDYCE, M., JOHNSON, K. E. & HART, L. G. 2004. The migration of physicians from sub-Saharan Africa to the United States of

America: measures of the African brain drain. *Human Resources for Health* [Online], 2. Available: http://www.human-resources-health.com/content/2/1/17 [Accessed 20 October 2014].

HAMDY, H. 2006. Blueprinting for the assessment of health care professionals. *The Clinical Teacher,* 3**,** 175-179.

HARDEN, R. 2001. AMEE Guide No. 21: Curriculum mapping: a tool for transparent and authentic teaching and learning. *Medical Teacher,* 23**,** 123-137.

HARDEN, R. & CROSBY, J. 2000. AMEE Guide No 20: The good teacher is more than a lecturer-the twelve roles of the teacher. *Medical Teacher,* 22**,** 334-347.

HARDEN, R., STEVENSON, M., DOWNIE, W. W. & WILSON, G. 1975. Assessment of clinical competence using objective structured examination. *British Medical Journal,* 1**,** 447-451.

HARDEN, R. M. & CAIRNCROSS, R. G. 1980. Assessment of practical skills: The objective structured practical examination (OSPE). *Studies in Higher Education,* 5**,** 187-196.

HARDEN, R. M. & GLEESON, F. 1979. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education,* 13**,** 39-54.

HARDEN, R. M., GRANT, J., BUCKLEY, G. & HART, I. 1999. BEME Guide No. 1: Best evidence medical education. *Medical Teacher,* 21**,** 553-562.

HARPER COLLINS 2014. Collins Dictionary. 12th ed. Glasgow: Harper Collins UK.

HARRISON, C. J., KÖNINGS, K. D., SCHUWIRTH, L. W., WASS, V. & VAN DER VLEUTEN, C. P. 2017. Changing the culture of assessment: the dominance of the summative assessment paradigm. *BMC Medical Education,* 17**,** 73.

HARTMAN, S. L. & NELSON, M. S. 1992. What we say and what we do: Self-reported teaching behavior versus performances in written simulations among medical school faculty. *Academic Medicine,* 67**,** 522-527.

HAUER, K. E., HODGSON, C. S., KERR, K. M., TEHERANI, A. & IRBY, D. M. 2005. A national study of medical student clinical skills assessment. *Academic Medicine,* 80**,** S25-S29.

HAWKINS, R. E., WELCHER, C. M., HOLMBOE, E. S., KIRK, L. M., NORCINI, J. J., SIMONS, K. B. & SKOCHELAK, S. E. 2015. Implementation of competency-based medical education: are we addressing the concerns and challenges? *Medical Education,* 49**,** 1086-1102.

HAYS, R. 2008. Assessment in medical education: roles for clinical teachers. *The Clinical Teacher,* 5**,** 23-27.

HEALTH PROFESSIONS COUNCIL OF SOUTH AFRICA. 2014. Core competencies for undergraduate students in clinical associate, dentistry and medical teaching and learning programmes in South Africa. Available: http://www.hpcsa.co.za/PBMedicalDental/Education [Accessed 02 April 2015].

HEALTH PROFESSIONS COUNCIL OF SOUTH AFRICA. 2017. *HPCSA Overview* [Online]. Pretoria, RSA: Health Professions Council of South Africa. Available: http://www.hpcsa.co.za/ [Accessed 25 July 2017 2017].

HENNING, E., VAN RENSBURG, W. & SMIT, B. 2004. *Finding your way in qualitative research,* Pretoria, Van Schaik.

HIFT, R. J. 2014. Should essays and other "open-ended"-type questions retain a place in written summative assessment in clinical medicine? *BMC Medical Education,* 14**,** 1.

HIGGS, J. & MCALLISTER, L. 2007. Educating clinical educators: using a model of the experience of being a clinical educator. *Medical Teacher,* 29**,** e51-e57.

HILL, F., KENDALL, K., GALBRAITH, K. & CROSSLEY, J. 2009. Implementing the undergraduate mini-CEX: a tailored approach at Southampton University. *Medical Education,* 43**,** 326-334.

HODGES, B. 2006. Medical education and the maintenance of incompetence. *Medical Teacher,* 28**,** 690-696.

HODGES, B., HANSON, M., MCNAUGHTON, N. & REGEHR, G. 2002. Creating, monitoring, and improving a psychiatry OSCE. *Academic Psychiatry,* 26**,** 134-161.

HODGES, B. D. 2010. A tea-steeping or i-Doc model for medical education? *Academic Medicine,* 85**,** S34-S44.

HOFFMANN, T. 1999. The meanings of competency. *Journal of European Industrial Training,* 23**,** 275-285.

HOLMBOE, E. S., SHERBINO, J., LONG, D. M., SWING, S. R. & FRANK, J. R. 2010. The role of assessment in competency-based medical education. *Medical Teacher,* 32**,** 676-682.

HOUGHTON, C., CASEY, D., SHAW, D. & MURPHY, K. 2013. Rigour in qualitative case-study research. *Nurse Researcher,* 20**,** 12-17.

HOWE, A., CAMPION, P., SEARLE, J. & SMITH, H. 2004. New perspectives - approaches to medical education at four new UK medical schools. *British Medical Journal,* 329**,** 327-331.

HSIEH, H.-F. & SHANNON, S. E. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research,* 15**,** 1277-1288.

HUDDLE, T. S. & HEUDEBERT, G. R. 2007. Viewpoint: Taking apart the art: The risk of anatomizing clinical competence. *Academic Medicine,* 82**,** 536-541.

HUDSON, J. N. & TONKIN, A. L. 2004. Evaluating the impact of moving from discipline-based to integrated assessment. *Medical Education,* 38**,** 832-843.

HULME, D. 2009. The Millennium Development Goals (MDGs): a short history of the world's biggest promise. Available: http://www.bwpi.manchester.ac.uk/medialibrary/publications/working_papers/bwpi-wp-10009.pdf [Accessed 01 October 2014].

HUMPHREY-MURTO, S., SMEE, S., TOUCHIE, C., WOOD, T. J. & BLACKMORE, D. E. 2005. A comparison of physician examiners and trained assessors in a high-stakes OSCE setting. *Academic Medicine,* 80**,** S59-S62.

INGHAM, A. 2011. The great wall of medical school: a comparison of barrier examinations across Australian medical schools. *Aust Med Student J* [Online], 2. Available: http://issuu.com/amsj/docs/amsj_v2_i2?e=1622759/2803784 [Accessed 13 April 2014].

INSTITUTE FOR INTERNATIONAL MEDICAL EDUCATION 2006. IIME Database of Medical Schools. July 2006 ed. New York.

ISSENBERG, S. B. 2011. Ottawa 2010 Conference - Consensus Statements and Recommendations. *Medical Teacher,* 33**,** 181-182.

JANSEN, J. D. 1998. Curriculum reform in South Africa: A critical analysis of outcomes-based education. *Cambridge Journal of Education,* 28**,** 321-331.

JEFFERIES, A., SIMMONS, B. & REGEHR, G. 2007. The effect of candidate familiarity on examiner OSCE scores. *Medical Education,* 41**,** 888-891.

JELOVSEK, J. E., KOW, N. & DIWADKAR, G. B. 2013. Tools for the direct observation and assessment of psychomotor skills in medical trainees: a systematic review. *Medical Education,* 47**,** 650-673.

JERIS, L., JOHNSON, K., ISOPAHKALA, U., WINTERTON, J. & ANTHONY, K. 2005. The politics of competence: Views from around the globe. *Human Resource Development International,* 8**,** 379-384.

JHA, V., BEKKER, H., DUFFY, S. & ROBERTS, T. 2006. Perceptions of professionalism in medicine: a qualitative study. *Medical Education,* 40**,** 1027-1036.

JOHNSTON, J. L., LUNDY, G., MCCULLOUGH, M. & GORMLEY, G. J. 2013. The view from over there: reframing the OSCE through the experience of standardised patient raters. *Medical Education,* 47**,** 899-909.

KARLE, H. 2008. International recognition of basic medical education programmes. *Medical Education,* 42**,** 12-17.

KASSEBAUM, N. J., BERTOZZI-VILLA, A., COGGESHALL, M. S., SHACKELFORD, K. A., STEINER, C., HEUTON, K. R., GONZALEZ-MEDINA, D., BARBER, R., HUYNH, C., DICKER, D., TEMPLIN, T., WOLOCK, T. M., ABBASOGLU OZGOREN, A., ABD-ALLAH, F., ABERA, S. F., ACHOKI, T., ADELEKAN, A., ADEMI, Z., ADOU, A. K., ADSUAR, J. C., AGARDH, E. E., AKENA, D., ALASFOOR, D., ALEMU, Z. A., ALFONSO-CRISTANCHO, R., ALHABIB, S., ALI, R., AL KAHBOURI, M. J., ALLA, F., ALLEN, P. J., ALMAZROA, M. A., ALSHARIF, U., ALVAREZ, E., ALVIS-GUZMAN, N., AMANKWAA, A. A., AMARE, A. T., AMINI, H., AMMAR, W., ANTONIO, C. A. T., ANWARI, P., ARNLOV, J., ARSENIJEVIC, V. S. A., ARTAMAN, A., ASAD, M. M., ASHGAR, R. J., ASSADI, R., ATKINS, L. S., BADAWI, A., BALAKRISHNAN, K., BASU, A., BASU, S., BEARDSLEY, J., BEDI, N., BEKELE, T., BELL, M. L., BERNABE, E., BEYENE, T. J., BHUTTA, Z., BIN ABDULHAK, A., BLORE, J., BORA BASARA, B., BOSE, D., BREITBORDE, N., CARDENAS, R., CASTANEDA-ORJUELA, C. A., CASTRO, R. E., CATALA-LOPEZ, F., CAVLIN, A., CHANG, J.-C., CHE, X., CHRISTOPHI, C. A., CHUGH, S. S., CIRILLO, M., COLQUHOUN, S. M., COOPER, L. T., COOPER, C., DA COSTA LEITE, I., DANDONA, L., DANDONA, R., DAVIS, A., DAYAMA, A., DEGENHARDT, L., DE LEO, D., DEL POZO-CRUZ, B., DERIBE, K., DESSALEGN, M., DEVEBER, G. A., DHARMARATNE, S. D., DILMEN, U., DING, E. L., DORRINGTON, R. E., DRISCOLL, T. R., ERNAKOV, S. P., ESTEGHAMATI, A., FARAON, E. J. A., FARZADFAR, F., FELICIO, M. M., FERESHTEHNEJAD, S.-M., FERREIRA DE LIMA, G. M., FOROUNZANFAR, M. H., et al. 2014. Global, regional, and national levels and causes of maternal mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet,* 384**,** 980-1004.

KEARNEY, R. A., PUCHALSKI, S. A., YANG, H. Y. & SKAKUN, E. N. 2002. The inter-rater and intra-rater reliability of a new Canadian oral examination format in anesthesia is fair to good. *Canadian Journal of Anesthesia,* 49**,** 232-236.

KERR, S., MULLER, D., MCKINON, W. & MC INERNEY, P. 2016. An online formative assessment tool to prepare students for summative assessment in physiology. *African Journal of Health Professions Education,* 8**,** 72-76.

KINFU, Y., DAL POZ, M. R., MERCER, H. & EVANS, D. B. 2009. The health worker shortage in Africa: are enough physicians and nurses being trained? *Bulletin of the World Health Organization,* 87**,** 225-230.

KNIGHT, P. 2006. The local practices of assessment. *Assessment & Evaluation in Higher Education,* 31**,** 435-452.

KNIGHT, P. T. 2002. Summative assessment in higher education: practices in disarray. *Studies in Higher Education,* 27**,** 275-286.

KOGAN, J. R., HOLMBOE, E. S. & HAUER, K. E. 2009. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA,* 302**,** 1316-1326.

KREFTING, L. 1991. Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy,* 45**,** 214-222.

KUPER, A., LINGARD, L. & LEVINSON, W. 2008. Critically appraising qualitative research. *British Medical Journal,* 337**,** 687-689.
KYBURZ-GRABER, R. 2004. Does case-study methodology lack rigour? The need for quality criteria for sound case-study research, as illustrated by a recent case in secondary and higher education. *Environmental Education Research,* 10**,** 53-65.
LABAREE, D. F. 1998. Educational researchers: Living with a lesser form of knowledge. *Educational Researcher,* 27**,** 4-12.
LE DEIST, F. D. & WINTERTON, J. 2005. What is competence? *Human Resource Development International,* 8**,** 27-46.
LESHEM, S. & TRAFFORD, V. 2007. Overlooking the conceptual framework. *Innovations in Education and Teaching International,* 44**,** 93-105.
LEUNG, W.-C. 2002. Competency based medical training: review. *British Medical Journal,* 325**,** 693-696.
LEWIS, J. 2003. Design issues. *In:* RITCHIE, J. & LEWIS, J. (eds.) *Qualitative Research Practice: A guide for social science students and researchers.* London: Sage.
LIAISON COMMITTEE ON MEDICAL EDUCATION. 2013. *Overview: Accreditation and the LCME* [Online]. Chicago IL & Washington DC: Liaison Committee on Medical Education. Available: http://www.lcme.org/overview.htm [Accessed 05 August 2013].
LIBRARY OF CONGRESS. 2010. *African and Middle Eastern Reading Room - List of Sub-Saharan African Countries* [Online]. Available: https://www.loc.gov/rr/amed/guide/afr-countrylist.html [Accessed 09 October 2014].
MAKERERE UNIVERSITY. 2015a. *Makerere University College of Health Sciences profile page* [Online]. Available: http://chs.mak.ac.ug/content/college-profile [Accessed 02 July 2015].
MAKERERE UNIVERSITY. 2015b. *Makerere University historical background* [Online]. Available: http://mak.ac.ug/about-makerere/historical-background [Accessed 02 July 2015].
MALIK, A. S. & MALIK, R. H. 2011. Twelve tips for developing an integrated curriculum. *Medical Teacher,* 33**,** 99-104.
MALONE, K. & SUPRI, S. 2012. A critical time for medical education: the perils of competence-based reform of the curriculum. *Advances in Health Sciences Education,* 17**,** 241-246.
MALTERUD, K. 2001a. The art and science of clinical knowledge: evidence beyond measures and numbers. *The Lancet,* 358**,** 397-400.
MALTERUD, K. 2001b. Qualitative research: standards, challenges, and guidelines. *The Lancet,* 358**,** 483-488.
MATHISON, S. 1988. Why triangulate? *Educational Researcher,* 17**,** 13-17.
MAVIS, B. E., COLE, B. L. & HOPPE, R. B. 2001. A survey of student assessment in US medical schools: the balance of breadth versus fidelity. *Teaching and Learning in Medicine,* 13**,** 74-79.
MAYS, N. & POPE, C. 1995. Qualitative research: Rigour and qualitative research. *British Medical Journal,* 311**,** 109-112.
MAYS, N. & POPE, C. 2000. Qualitative research in health care: Assessing quality in qualitative research. *British Medical Journal,* 320**,** 50-52.
MCCOUBRIE, P. 2004. Improving the fairness of multiple-choice questions: a literature review. *Medical Teacher,* 26**,** 709-712.
MCCRORIE, P. & BOURSICOT, K. A. 2009. Variations in medical school graduating examinations in the United Kingdom: Are clinical competence standards comparable? *Medical Teacher,* 31**,** 223-229.

MCGLOIN, S. 2008. The trustworthiness of case study methodology. *Nurse Researcher,* 16**,** 45-55.
MCKINLEY, D. W. & NORCINI, J. J. 2014. How to set standards on performance-based examinations: AMEE Guide No. 85. *Medical Teacher,* 36**,** 97-110.
MCLEAN, M., CILLIERS, F. & VAN WYK, J. M. 2008. Faculty development: Yesterday, today and tomorrow. *Medical Teacher,* 30**,** 555-584.
MCMULLAN, M., ENDACOTT, R., GRAY, M. A., JASPER, M., MILLER, C. M., SCHOLES, J. & WEBB, C. 2003. Portfolios and assessment of competence: a review of the literature. *Journal of Advanced Nursing,* 41**,** 283-294.
MEDICAL EDUCATION PARTNERSHIP INITIATIVE. 2013. *Medical Education Partnership Initiative Summary Report 2010-2012: Building capacity for change* [Online]. Available: http://www.mepinetwork.org/ [Accessed 26 October 2013].
MEMON, M. A., JOUGHIN, G. R. & MEMON, B. 2010. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Advances in Health Sciences Education,* 15**,** 277-289.
MERRIAM WEBSTER 2015. Merriam-Webster Dictionary. Merriam-Webster.
MILES, M. B. & HUBERMAN, A. M. 1994. *Qualitative data analysis: An expanded sourcebook,* Thousand Oaks, CA, Sage.
MILLER, B. M., EICHBAUM, Q., BRADY, D. W. & MOORE, D. E. 2011. Aligning health sciences education with health needs in developing countries. *Academic Medicine,* 86**,** e10.
MILLER, G. E. 1990. The assessment of clinical skills/competence/performance. *Academic Medicine,* 65**,** S63-S67.
MONEKOSSO, G. 2014. A brief history of medical education in Sub-Saharan Africa. *Academic Medicine,* 89**,** S11-S15.
MOORE, D. R., CHENG, M.-I. & DAINTY, A. R. 2002. Competence, competency and competencies: performance assessment in organisations. *Work Study,* 51**,** 314-319.
MORCKE, A. M., DORNAN, T. & EIKA, B. 2013. Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Advances in Health Sciences Education,* 18**,** 851-863.
MORRISON, J. & WATT, G. 2002. External examining–a challenge for new curricula? *Medical Education,* 36**,** 308-309.
MUHR, T. 1997. ATLAS. ti. *Berlin: Scientific Software Development*.
MULLAN, F. & FREHYWOT, S. 2008. Non-physician clinicians in 47 sub-Saharan African countries. *The Lancet,* 370**,** 2158-2163.
MULLAN, F., FREHYWOT, S., OMASWA, F., BUCH, E., CHEN, C., GREYSEN, S. R., WASSERMANN, T., ABUBAKR, D. E. E., AWASES, M., BOELEN, C., DIOMANDE, M.-M., DOVLO, D., FERRO, J., HAILEAMLAK, A., IPUTO, J., JACOBS, M., KARIM, K. A., MIPANDO, M., MONEKOSSO, G. L., OLAPADE-OLAOPA, E. O., RUGARABAMU, P., SEWANKAMBO, N. K., ROSS, H., AYAS, H., CHALE, S. B., SOEURETTE, C., COHEN, J., HAILE-MARIAM, T., HAMBURGER, E., JOLLEY, L., KOLARS, J. C., KOMBE, G. & NEUSY, A.-J. 2011. Medical schools in sub-Saharan Africa. *The Lancet,* 377**,** 1113-1121.
MULLAN, F., FREHYWOT, S., OMASWA, F., SEWANKAMBO, N., TALIB, Z., CHEN, C., KIARIE, J. & KIGULI-MALWADDE, E. 2012. The Medical Education Partnership Initiative: PEPFAR's effort to boost health worker education to strengthen health systems. *Health Affairs,* 31**,** 1561-1572.
MURRAY, C. J., ORTBLAD, K. F., GUINOVART, C., LIM, S. S., WOLOCK, T. M., ROBERTS, D. A., DANSEREAU, E. A., GRAETZ, N., BARBER, R. M., BROWN, J. C., WANG, H., DUBER, H. C., NAGHAVI, M., DICKER, D., DANDONA, L.,

SALOMON, J. A., HEUTON, K. R., FOREMAN, K., PHILLIPS, D. E., FLEMING, T. D., FLAXMAN, A. D., PHILLIPS, B. K., JOHNSON, E. K., COGGESHALL, M. S., ABD-ALLAH, F., ABERA, S. F., ABRAHAM, J. P., ABUBAKAR, I., ABU-RADDAD, L. J., ABU-RMEILEH, N., ACHOKI, T., ADEYEMO, A. O., ADOU, A. K., ADSUARI, J. C., AGARDH, E. E., AKENA, D., AL KAHBOURI, M. J., ALASFOOR, D., ALBITTAR, M. I., ALCALA-CERRA, G., ALEGRETTI, M. A., ALEMU, Z. A., ALFONSO-CRISTANCHO, R., ALHABIB, S., ALI, R., ALLA, F., ALLEN, P. J., ALSHARIF, U., ALVAREZ, E., ALVIS-GUZMAN, N., AMANKWAA, A. A., AMARE, A. T., AMINI, H., AMMAR, W., ANDERSON, B. O., ANTONIO, C. A., ANWARI, P., ARNLOV, J., ARSENIJEVIC, V. S. A., ARTAMAN, A., ASHGARI, R. J., ASSADI, R., ATKINS, L. S., BADAWI, A., BALAKRISHNAN, K., BANERJEE, A., BASU, S., BEARDSLEY, J., BEKELE, T., BELL, M. L., BERNABE, E., BEYENE, T. J., BHALA, N., BHALLA, A., BHUTTA, Z. A., BIN ABDULHAK, A., BINAGWAHO, A., BLORE, J. D., BASARA, B. B., BOSE, D., BRAININ, M., BREITBORDE, N., CASTANEDA-ORJUELA, C. A. & CATALA-LOPEZ, F. 2014. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet,* 384**,** 1005-1070.

MURRAY, C. J. L. & LOPEZ, A. D. 1997. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *The Lancet,* 349**,** 1269-1276.

MURRAY, C. J. L., VOS, T., LOZANO, R., NAGHAVI, M., FLAXMAN, A. D., MICHAUD, C., EZZATI, M., SHIBUYA, K., SALOMON, J. A., ABDALLA, S., ABOYANS, V., ABRAHAM, J., ACKERMAN, I., AGGARWAL, R., AHN, S. Y., ALI, M. K., ALMAZROA, M., ALVARADO, M., ANDERSON, H. R., ANDERSON, L. M., ANDREWS, K. G., ATKINSON, C., BADDOUR, L. M., BAHALIM, A. N., BARKER-COLLO, S., BARRERO, L. H., BARTLES, D. H., BASANEZ, M.-G., BAXTER, A., BELL, M. L., BENJAMIN, E. J., BENNETT, D., BERNABE, E., BHALLA, K., BHANDARI, B., BIKBOV, B., BIN ABDULHAK, A., BIRBECK, G., BLACK, J. A., BLENCOWE, H., BLORE, J. D., BLYTH, F., BOLLIGER, I., BONAVENTURE, A., BOUFOUS, S., BOURNE, R., BOUSSINESQ, M., BRAITHWAITE, T., BRAYNE, C., BRIDGETT, L., BROOKER, S., BROOKS, P., BRUGHA, T. S., BRYAN-HANCOCK, C., BUCELLO, C., BUCHBINDER, R., BUCKLE, G., BUDKE, C. M., BURCH, M., BURNEY, P., BURSTEIN, R., CALABRIA, B., CAMPBELL, B., CANTER, C. E., CARABIN, H., CARAPETIS, J., CARMONA, L., CELLA, C., CHARLSON, F., CHEN, H., CHENG, A. T.-A., CHOU, D., CHUGH, S. S., COFFENG, L. E., COLAN, S. D., COLQUHOUN, S., COLSON, K. E., CONDON, J., CONNOR, M. D., COOPER, L. T., CORRIERE, M., CORTINOVIS, M., COURVILLE DE VACCARO, K., COUSER, W., COWIE, B. C., CRIQUI, M. H., CROSS, M., DABHADKAR, K. C., DAHIYA, M., DAHODWALA, N., DAMSERE-DERRY, J., DANAEI, G., DAVIS, A., DE LEO, D., DEGENHARDT, L., DELLAVALLE, R., DELOSSANTOS, A., DENENBERG, J., DERRETT, S., DES JARLAIS, D. C., et al. 2012. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet,* 380**,** 2197-2223.

NATRIELLO, G. 1987. The impact of evaluation processes on students. *Educational Psychologist,* 22**,** 155-175.

NEWBLE, D. I. 1992. Assessing clinical competence at the undergraduate level. *Medical Education,* 26**,** 503-511.

NEWTON, P. E. 2007. Clarifying the purposes of educational assessment. *Assessment in Education,* 14**,** 149-170.
NIEHAUS, D. J. H., JORDAAN, E., KOEN, L., MASHILE, M. & MALL, S. 2012. Applicability and fairness of the oral examination in undergraduate psychiatry training in South Africa. *African Journal of Psychiatry,* 15**,** 119-123.
NORCINI, J., ANDERSON, B., BOLLELA, V., BURCH, V., COSTA, M. J., DUVIVIER, R., GALBRAITH, R., HAYS, R., KENT, A. & PERROTT, V. 2011. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher,* 33**,** 206-214.
NORCINI, J. & BURCH, V. 2007. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher,* 29**,** 855-871.
NORCINI, J., SWANSON, D., GROSSO, L. & WEBSTER, G. 1985. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. *Medical Education,* 19**,** 238-247.
NORCINI, J. J. 2002. The death of the long case? *British Medical Journal,* 324**,** 408-409.
NORCINI, J. J. 2003a. ABC of learning and teaching in medicine: Work based assessment. *BMJ: British Medical Journal,* 326**,** 753-755.
NORCINI, J. J. 2003b. Setting standards on educational tests. *Medical Education,* 37**,** 464-469.
NORCINI, J. J. 2003c. Work based assessment. *British Medical Journal,* 326**,** 753-755.
NORCINI, J. J., BLANK, L. L., DUFFY, F. D. & FORTNA, G. S. 2003. The mini-CEX: a method for assessing clinical skills. *Annals of Internal Medicine,* 138**,** 476-481.
NORCINI, J. J., LIPNER, R. S. & GROSSO, L. J. 2013. Assessment in the context of licensure and certification. *Teaching and Learning in Medicine,* 25**,** S62-S67.
NORCINI, J. J. & MCKINLEY, D. W. 2007. Assessment methods in medical education. *Teaching and Teacher Education,* 23**,** 239-250.
NORCINI, J. J. & SHEA, J. A. 1997. The credibility and comparability of standards. *Applied Measurement in Education,* 10**,** 39-59.
NORMAN, G. 2000. Examining the examination: Canadian versus US radiology certification exam. *Canadian Association of Radiologist Journal,* 51**,** 208-209.
NORMAN, G. 2006. Editorial - Outcomes, Objectives, and the Seductive Appeal of Simple Solutions. *Advances in Health Sciences Education,* 11**,** 217-220.
O'SULLIVAN, H., VAN MOOK, W., FEWTRELL, R. & WASS, V. 2012. Integrating professionalism into the curriculum: AMEE Guide No. 61. *Medical Teacher,* 34**,** e64-e77.
ODOI ADOME, R. & KITUTU, F. 2008. Creating an OSCE/OSPE in a resource-limited setting. *Medical Education,* 42**,** 525-526.
OLAPADE-OLAOPA, E. O., BAIRD, S., KIGULI-MALWADDE, E. & KOLARS, J. C. 2014. Growing partnerships: leveraging the power of collaboration through the Medical Education Partnership Initiative. *Academic Medicine,* 89**,** S19-S23.
OMER, M. I. A. 2005. Can medical education rise to the challenge of the African crisis? *Annals of Tropical Paediatrics: International Child Health,* 25**,** 227-241.
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. 2013. *OECD Health data 2013 - Frequently requested data* [Online]. Paris: Organisation for Economic Co-operation and Development Available: http://www.oecd.org/health/health-systems/oecdhealthdata2013-frequentlyrequesteddata.htm [Accessed 05 August 2013].
OTTAWA. 2015. *Ottawa Conference on the Assessment of Competence in Medicine and the Healthcare Professions* [Online]. Dundee: Association for Medical Education in

Europe. Available: http://www.amee.org/conferences/ottawa-conferences [Accessed 13 January 2015].
OVIASU, E. 2008. Clinical skills evaluation in a resource-constrained environment. *Medical Education,* 42**,** 524-525.
OXFORD UNIVERSITY PRESS 2015. Oxford Advanced Learners' Dictionary. 9th ed. Oxford: Oxford University Press.
PAGE, G., BORDAGE, G. & ALLEN, T. 1995. Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine,* 70**,** 194-201.
PALMER, E. J. & DEVITT, P. G. 2007. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Medical Education,* 7**,** 49.
PALMER, E. J., DUGGAN, P., DEVITT, P. G. & RUSSELL, R. 2010. The modified essay question: its exit from the exit examination? *Medical Teacher,* 32**,** e300-e307.
PANGARO, L. & TEN CATE, O. 2013. Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Medical Teacher,* 35**,** e1197-e1210.
PATRÍCIO, M. F., JULIÃO, M., FARELEIRA, F. & CARNEIRO, A. V. 2013. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Medical Teacher,* 35**,** 503-514.
PELL, G., FULLER, R., HOMER, M. & ROBERTS, T. 2010. How to measure the quality of the OSCE: A review of metrics–AMEE guide no. 49. *Medical Teacher,* 32**,** 802-811.
PLOWRIGHT, D. 2011. *Using mixed methods: Frameworks for an integrated methodology,* London, Sage.
PONNAMPERUMA, G. G., KARUNATHILAKE, I. M., MCALEER, S. & DAVIS, M. H. 2009. The long case and its modifications: a literature review. *Medical Education,* 43**,** 936-941.
POPE, C., ZIEBLAND, S. & MAYS, N. 2000. Qualitative research in health care: Analysing qualitative data. *British Medical Journal,* 320**,** 114-116.
POPHAM, W. J. 2011. Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator,* 46**,** 265-273.
PRIDEAUX, D. 2003. Curriculum design. *British Medical Journal,* 326**,** 268-270.
PRIDEAUX, D. & BLIGH, J. 2002. Research in medical education: asking the right questions. *Medical Education,* 36**,** 1114-1115.
QUALITY ASSURANCE AGENCY FOR HIGHER EDUCATION UK. 2017. *Chapter B7: External Examiners* [Online]. Gloucester: Quality Assurance Agency for Higher Education. Available: http://www.qaa.ac.uk/assuring-standards-and-quality/the-quality-code [Accessed 26 June 2017].
REES, C. E. 2004. The problem with outcomes-based curricula in medical education: insights from educational theory. *Medical Education,* 38**,** 593-598.
REES, C. E. & KNIGHT, L. V. 2007. Viewpoint: The trouble with assessing students' professionalism: Theoretical insights from sociocognitive psychology. *Academic Medicine,* 82**,** 46-50.
RETHANS, J. J., NORCINI, J., BARON-MALDONADO, M., BLACKMORE, D., JOLLY, B., LADUCA, T., LEW, S., PAGE, G. & SOUTHGATE, L. 2002. The relationship between competence and performance: implications for assessing practice performance. *Medical Education,* 36**,** 901-909.
REZNICK, R. K., SMEE, S., BAUMBER, J. S., COHEN, R., ROTHMAN, A., BLACKMORE, D. & BERARD, M. 1993. Guidelines for estimating the real cost of an objective structured clinical examination. *Academic Medicine,* 68**,** 513-7.

RITCHIE, J. 2003. The applications of qualitative methods to social research. *In:* RITCHIE, J. & LEWIS, J. (eds.) *Qualitative research practice: A guide for social science students and researchers.* London: Sage.

RITCHIE, J., LEWIS, J. & ELAM, G. 2003. Designing and selecting samples. *In:* RITCHIE, J. & LEWIS, J. (eds.) *Qualitative research practice: A guide for social science students and researchers.* London: Sage.

ROBERTS, C., NEWBLE, D., JOLLY, B., REED, M. & HAMPTON, K. 2006. Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Medical Teacher,* 28**,** 535-543.

ROBERTS, C., WASS, V., JONES, R., SARANGI, S. & GILLETT, A. 2003. A discourse analysis study of 'good'and 'poor'communication in an OSCE: a proposed new framework for teaching students. *Medical Education,* 37**,** 192-201.

ROTGANS, J. I. 2012. The themes, institutions, and people of medical education research 1988–2010: content analysis of abstracts from six journals. *Advances in Health Sciences Education,* 17**,** 515-527.

ROWE, C. 1995. Clarifying the use of competence and competency models in recruitment, assessment and staff development. *Industrial and Commercial Training,* 27**,** 12-17.

ROYAL COLLEGE OF PHYSICIANS AND SURGEONS OF CANADA WORKING GROUP 2000. CanMEDS 2000: Extract from the CanMEDS 2000 Project Societal Needs Working Group Report. *Medical Teacher,* 22**,** 549-554.

SACHS, J. D. 2012. From millennium development goals to sustainable development goals. *The Lancet,* 379**,** 2206-2211.

SADLER, D. R. 1989. Formative assessment and the design of instructional systems. *Instructional Science,* 18**,** 119-144.

SALDAÑA, J. 2013. *The coding manual for qualitative researchers,* Thousand Oaks, California, Sage.

SANDILANDS, D., GOTZMANN, A., ROY, M., ZUMBO, B. D. & DE CHAMPLAIN, A. 2014. Weighting checklist items and station components on a large-scale OSCE: Is it worth the effort? *Medical Teacher,* 36**,** 585-590.

SCHUWIRTH, L. W. & VAN DER VLEUTEN, C. P. 2003a. ABC of learning and teaching in medicine: Written assessment. *British Medical Journal,* 326**,** 643-645.

SCHUWIRTH, L. W. & VAN DER VLEUTEN, C. P. 2003b. The use of clinical simulations in assessment. *Medical Education,* 37**,** 65-71.

SCHUWIRTH, L. W. & VAN DER VLEUTEN, C. P. 2004. Different written assessment methods: what can be said about their strengths and weaknesses? *Medical Education,* 38**,** 974-979.

SCHUWIRTH, L. W. & VAN DER VLEUTEN, C. P. 2010. How to design a useful test: the principles of assessment. *In:* SWANWICK, T. (ed.) *Understanding medical education: evidence, theory and practice.* 2nd ed. Chichester: ASME in association with Wiley & Sons.

SCHUWIRTH, L. W. & VAN DER VLEUTEN, C. P. 2011. Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher,* 33**,** 478-485.

SCRIVEN, M. S. 1967. The methodology of evaluation (Perspectives of Curriculum Evaluation, and AERA monograph Series on Curriculum Evaluation, No. 1). *Chicago: Rand NcNally*.

SEARLE, J. 2004. Graduate entry medicine: what it is and what it isn't. *Medical Education,* 38**,** 1130-1132.

SEFTON, A. J. 2004. Graduate entry to medical school. *Medical Education,* 38**,** 1132-1134.

SHUMWAY, J. & HARDEN, R. 2003. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Medical Teacher,* 25**,** 569-584.

SIMOENS, S. & HURST, J. 2006. The Supply of Physician Services in OECD Countries. *OECD Health Working Papers, No. 21* [Online]. Available: http://dx.doi.org/10.1787/608402211700 [Accessed 09 April 2015].

SIMPSON, J. G., FURNACE, J., CROSBY, J., CUMMING, A. D., EVANS, P. A., FRIEDMAN BEN-DAVID, M., HARDEN, R. M., LLOYD, D., MCKENZIE, H., MCLACHLAN, J. C., MCPHATE, G. F., PERCY-ROBB, I. W. & MACPHERSON, S. G. 2002. The Scottish doctor - learning outcomes for the medical undergraduate in Scotland: A foundation for competent and reflective practitioners. *Medical Teacher,* 24**,** 136-143.

STEINERT, Y., CRUESS, S., CRUESS, R. & SNELL, L. 2005. Faculty development for teaching and evaluating professionalism: from programme design to curriculum change. *Medical Education,* 39**,** 127-136.

STEINERT, Y., MANN, K., CENTENO, A., DOLMANS, D., SPENCER, J., GELULA, M. & PRIDEAUX, D. 2006. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical Teacher,* 28**,** 497-526.

STEINERT, Y. & MANN, K. V. 2006. Faculty development: principles and practices. *Journal of Veterinary Medical Education,* 33**,** 317-324.

STEINERT, Y., NAISMITH, L. & MANN, K. 2012. Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19. *Medical Teacher,* 34**,** 483-503.

STELLENBOSCH UNIVERSITY. 2015a. *Stellenbosch University historical background* [Online]. Available: http://www.sun.ac.za/english/about-us/historical-background [Accessed 02 July 2015].

STELLENBOSCH UNIVERSITY. 2015b. *Yearbook for the Faculty of Medicine and Health Sciences* [Online]. Available: http://www.sun.ac.za/english/Documents/Yearbooks/Current/Medicine.pdf#search=Search%2E%2E%2Eyear%20book [Accessed 02 July 2015].

STIGGINS, R. J. 1991. Assessment Literacy. *Phi Delta Kappan,* 72**,** 534-539.

STIGGINS, R. J. 1995. Assessment literacy for the 21st century. *Phi Delta Kappan,* 77**,** 238-245.

STUPART, D., GOLDBERG, P., KRIGE, J. & KHAN, D. 2008. Does examiner bias in undergraduate oral and clinical surgery examinations occur? *South African Medical Journal,* 98**,** 805-807.

SWANSON, D. B., NORMAN, G. R. & LINN, R. L. 1995. Performance-based assessment: Lessons from the health professions. *Educational Researcher,* 24**,** 5-11,35.

SWANWICK, T. 2008. See one, do one, then what? Faculty development in postgraduate medical education. *Postgraduate Medical Journal,* 84**,** 339-343.

TALBOT, M. 2004. Monkey see, monkey do: a critique of the competency model in graduate medical education. *Medical Education,* 38**,** 587-592.

TALIB, Z. M., BAINGANA, R. K., SAGAY, A. S., VAN SCHALKWYK, S. C., MEHTSUN, S. & KIGULI-MALWADDE, E. 2013. Investing in community-based education to improve the quality, quantity, and retention of physicians in three African countries. *Education for Health,* 26**,** 109-114.

TAN, C. P. L., VAN SCHALKWYK, S. C., BEZUIDENHOUT, J. & CILLIERS, F. J. 2016. Mapping undergraduate exit-level assessment in a medical programme: A blueprint for clinical competence? *African Journal of Health Professions Education,* 8**,** 45-49.

TARAS, M. 2005. Assessment - Summative and Formative - Some theoretical reflections. *British Journal of Educational Studies,* 53**,** 466-478.

TAVAKOL, M. & SANDARS, J. 2014a. Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part I. *Medical Teacher,* 36**,** 746-756.
TAVAKOL, M. & SANDARS, J. 2014b. Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part II. *Medical Teacher,* 36**,** 838-848.
TAYLOR, D. S. 1988. The meaning and use of the term "competence" in Linguistics and Applied Linguistics. *Applied Linguistics,* 9**,** 148-168.
THE COMMONWEALTH. 2015. *The Commonwealth* [Online]. Available: http://thecommonwealth.org/about-us [Accessed 21 September 2015].
THE COMMONWEALTH. 2017. *Botswana: History* [Online]. Available: http://thecommonwealth.org/our-member-countries/botswana/history [Accessed 21 May 2017].
THURMOND, V. A. 2001. The point of triangulation. *Journal of Nursing Scholarship,* 33**,** 253-258.
UN MILLENNIUM PROJECT 2005. Investing in Development: A Practical Plan to Achieve the Millennium Development Goals. 2005 ed. New York: United Nations Development Programme.
UNITED NATIONS. 2000. Millennium Declaration. Available: http://www.unsceb.org/content/united-nations-millennium-summit-6-8-september-2000.
UNITED NATIONS. 2001. Road Map Towards the Implementation of the United Nations Millennium Declaration: Report of the Secretary-General. Available: http://www.un.org/en/ga/search/view_doc.asp?symbol=A/56/326 [Accessed 27 October 2014].
UNITED NATIONS. 2008. *Official list of MDG indicators* [Online]. New York: United Nations. Available: http://mdgs.un.org/unsd/mdg/Host.aspx?Content=Indicators/OfficialList.htm [Accessed 11 April 2015].
UNITED NATIONS. 2013. Millennium Development Goals Report 2013. Available: http://www.un.org/millenniumgoals/pdf/report-2013/mdg-report-2013-english.pdf [Accessed 16 September 2013].
UNITED NATIONS. 2014a. *Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings* [Online]. New York: United Nations Statistics Division. Available: http://unstats.un.org/unsd/methods/m49/m49regin.htm#africa [Accessed 09 October 2014].
UNITED NATIONS. 2014b. Millennium Development Goals Report 2014. Available: http://mdgs.un.org/unsd/mdg/Resources/Static/Products/Progress2014/English2014.pdf [Accessed 09 October 2014].
UNITED NATIONS. 2014c. *United Nations Statistics: Pocketbook, 2014* [Online]. New York: United Nations. Available: http://unstats.un.org/unsd/publications/pocketbook [Accessed 09 October 2014].
UNITED NATIONS. 2016. *Sustainable Development Goals* [Online]. United Nations Available: http://www.un.org/sustainabledevelopment/ [Accessed 09 Dec 2016].
UNIVERSITY OF BOTSWANA. 2015a. *University of Botswana: Faculty of Health Sciences History* [Online]. Available: http://www.ub.bw/content/id/1962/pid/1174/ac/1/fac/13/History/ [Accessed 02 July 2015].
UNIVERSITY OF BOTSWANA. 2015b. *University of Botswana: History* [Online]. Available: http://www.ub.bw/content/id/1366/History/ [Accessed 02 July 2015].

VAN DER VLEUTEN, C. 2000. Validity of final examinations in undergraduate medical training. *British Medical Journal,* 321**,** 1217-1219.

VAN DER VLEUTEN, C., DOLMANS, D. & SCHERPBIER, A. 2000. The need for evidence in education. *Medical Teacher,* 22**,** 246-250.

VAN DER VLEUTEN, C., SCHUWIRTH, L., SCHEELE, F., DRIESSEN, E. & HODGES, B. 2010. The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology,* 24**,** 703-719.

VAN DER VLEUTEN, C. P. 1996. The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education,* 1**,** 41-67.

VAN DER VLEUTEN, C. P. & SCHUWIRTH, L. W. 2005. Assessing professional competence: from methods to programmes. *Medical Education,* 39**,** 309-317.

VAN MOOK, W. N., GORTER, S. L., O'SULLIVAN, H., WASS, V., SCHUWIRTH, L. W. & VAN DER VLEUTEN, C. P. 2009. Approaches to professional behaviour assessment: Tools in the professionalism toolbox. *European Journal of Internal Medicine,* 20**,** e153-e157.

VAN SCHALKWYK, S., BEZUIDENHOUT, J., BURCH, V. C., CLARKE, M., CONRADIE, H., VAN HEERDEN, B. & DE VILLIERS, M. 2012. Developing an educational research framework for evaluating rural training of health professionals: A case for innovation. *Medical Teacher,* 34**,** 1064-1069.

VAN TARTWIJK, J. & DRIESSEN, E. W. 2009. Portfolios for assessment and learning: AMEE Guide no. 45. *Medical Teacher,* 31**,** 790-801.

VELOSKI, J. J., FIELDS, S. K., BOEX, J. R. & BLANK, L. L. 2005. Measuring Professionalism: A Review of Studies with Instruments Reported in the Literature between 1982 and 2002. *Academic Medicine,* 80**,** 366-370.

WALSH, K. 2013. When I say… triangulation. *Medical Education,* 47**,** 866.

WALUBO, A., BURCH, V., PARMAR, P., RAIDOO, D., CASSIMJEE, M., ONIA, R. & OFEI, F. 2003. A model for selecting assessment methods for evaluating medical students in African medical schools. *Academic Medicine,* 78**,** 899-906.

WANG, H., LIDDELL, C. A., COATES, M. M., MOONEY, M. D., LEVITZ, C. E., SCHUMACHER, A. E., APFEL, H., IANNARONE, M., PHILLIPS, B., LOFGREN, K. T., SANDAR, L., DORRINGTON, R. E., RAKOVAC, I., JACOBS, T. A., LIANG, X., ZHOU, M., ZHU, J., YANG, G., WANG, Y., LIU, S., YICHONG, L., OZGOREN, A. A., ABERA, S. F., ABUBAKAR, I., ACHOKI, T., ADELEKAN, A., ADEMI, Z., ALEMU, Z. A., ALLEN, P. J., ALMAZROA, M. A., ALVAREZ, E., AMANKWAA, A. A., AMARE, A. T., AMMART, W., ANWARI, P., ARGESEANU CUNNINGHAM, S., ASAD, M. M., ASSADI, R., BANERJEE, A., BASU, S., BEDI, N., BEKELE, T., BELL, M. L., BHUTTA, Z., BLORE, J., BASARA, B. B., BOUFOS, S., BREITBORDE, N., BRUCE, N. G., BUI, L. N., CARAPETIS, J. R., CARDENAS, R., CARPENTER, D. O., CASO, V., CASTRO, R. E., CATALA-LOPEZ, F., CAVLIN, A., CHE, X., CHIANG, P. P.-C., CHOWDHURY, R., CHRISTOPHI, C. A., CHUANG, T.-W., CIRILLO, M., DA COSTA LEITE, I., COURVILLE, K. J., DANDONA, L., DANDONA, R., DAVIS, A., DAYAMA, A., DERIBE, K., DHARMARATNE, S. D., DHERANI, M. K., DILMEN, U., DING, E. L., EDMOND, K. M., ERNAKOV, S. P., FARZADFAR, F., FERESHTEHNEJAD, S.-M., FIJABI, D. O., FOIGT, N., FOROUZANFAR, M. H., GARCIA, A. C., GELEIJNSE, J. M., GESSNER, B. D., GOGINASHVILI, K., GONA, P., GOTO, A., GOUDA, H. N., GREEN, M. A., GREENWELL, K. F., GUGNANI, H. C., GUPTA, R., HAMADEH, R. R., HAMMAMI, M., HARB, H. L.,

HAY, S., HEDAYATI, M. T., HOSGOOD, H. D., HOY, D. G., IDRISOV, B. T., et al. 2014. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet,* 384**,** 957-979.

WASS, V., JONES, R. & VAN DER VLEUTEN, C. 2001a. Standardized or real patients to test clinical competence? The long case revisited. *Medical Education,* 35**,** 321-325.

WASS, V., MCGIBBON, D. & VAN DER VLEUTEN, C. 2001b. Composite undergraduate clinical examinations: how should the components be combined to maximize reliability? *Medical Education,* 35**,** 326-330.

WASS, V. & VAN DER VLEUTEN, C. 2004. The long case. *Medical Education,* 38**,** 1176-1180.

WASS, V., VAN DER VLEUTEN, C., SHATZER, J. & JONES, R. 2001c. Assessment of clinical competence. *The Lancet,* 357**,** 945-949.

WASS, V., WAKEFORD, R., NEIGHBOUR, R. & VAN DER VLEUTEN, C. 2003. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Medical Education,* 37**,** 126-131.

WATSON, R., STIMPSON, A., TOPPING, A. & POROCK, D. 2002. Clinical competence assessment in nursing: a systematic review of the literature. *Journal of Advanced Nursing,* 39**,** 421-431.

WEAVER, K. & OLSON, J. K. 2006. Understanding paradigms used for nursing research. *Journal of Advanced Nursing,* 53**,** 459-469.

WEBBER, R. 1992. Structured short-answer questions: an alternative examination method. *Medical Education,* 26**,** 58-62.

WESTON, P. S. & SMITH, C. A. 2014. The use of mini-CEX in UK foundation training six years following its introduction: Lessons still to be learned and the benefit of formal teaching regarding its utility. *Medical Teacher,* 36**,** 155-163.

WILIAM, D. & BLACK, P. 1996. Meanings and Consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal,* 22**,** 537-548.

WILKERSON, L. & IRBY, D. M. 1998. Strategies for improving teaching practices: a comprehensive approach to faculty development. *Academic Medicine,* 73**,** 387-396.

WILKINSON, T. 2007. Assessment of clinical performance: gathering evidence. *Internal Medicine Journal,* 37**,** 631-636.

WILKINSON, T. J. & FRAMPTON, C. M. 2004. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Medical Education,* 38**,** 1111-1116.

WILKINSON, T. J., FRAMPTON, C. M., THOMPSON-FAWCETT, M. & EGAN, T. 2003. Objectivity in objective structured clinical examinations: Checklists are no substitute for examiner commitment. *Academic Medicine,* 78**,** 219-223.

WILKINSON, T. J., WADE, W. B. & KNOCK, L. D. 2009. A Blueprint to Assess Professionalism: Results of a Systematic Review. *Academic Medicine,* 84**,** 551-558.

WIUM, A. & DU PLESSIS, S. 2016. The usefulness of a tool to assess reflection in a service-learning experience. *African Journal of Health Professions Education,* 8**,** 178-183.

WOJTCZAK, A. 2002. Glossary of medical education terms: Part 1. *Medical Teacher,* 24**,** 216-219.

WORLD DIRECTORY OF MEDICAL SCHOOLS. 2014. *World Directory of Medical Schools* [Online]. World Federation for Medical Education and the Foundation for the

Advancement of International Medical Education. Available: http://www.wdoms.org/ [Accessed 27 April 2015].

WORLD HEALTH ORGANISATION. 2006. *Working Together for Health: The World Health Report 2006* [Online]. Geneva: World Health Organisation. Available: http://www.who.int/whr/2006/en/ [Accessed 15 July 2013].

WORLD HEALTH ORGANISATION. 2012. *Millennium Development Goals (MDGs). Fact sheet no. 290* [Online]. Geneva: World Health Organisation. Available: http://www.who.int/mediacentre/factsheets/fs290/en/ [Accessed 14 July 2013].

WORLD HEALTH ORGANISATION. 2014. *World Health Statistics 2014* [Online]. Geneva: World Health Organisation. Available: http://www.who.int/gho/publications/world_health_statistics/2014/en/ [Accessed 09 October 2014].

WORLD HEALTH ORGANISATION. 2015. *Metrics: Disability-Adjusted Life Year (DALY). Health statistics and information systems* [Online]. Geneva: World Health Organisation. Available: http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/ [Accessed 02 April 2015].

WORLD HEALTH ORGANIZATION. 2012. *A review of the work of World Health Organization in the African Region 2010-2011* [Online]. Luanda, Angola: World Health Organization. Available: http://www.afro.who.int/en/media-centre/pressreleases/item/5162-a-review-of-the-work-of-world-health-organization-in-the-african-region.html [Accessed 09 Dec 2016].

WORLD HEALTH ORGANIZATION. 2017. *Environmental health in emergencies: Vulnerable groups* [Online]. World Health Organisation. Available: http://www.who.int/environmental_health_emergencies/vulnerable_groups/en/ [Accessed 19 July 2017 2017].

WRIGHT, J. D. 2016. The Application of Simple, Practical and Affordable Standard Setting Methods in Small Groups of Students. *MedEdPublish*, 5.

YANHUA, C. & WATSON, R. 2011. A review of clinical competence assessment in nursing. *Nurse Education Today,* 31**,** 832-836.

YIN, R. K. 2014. *Case Study Research: Design and Methods,* Thousand Oaks, CA, Sage.